GESELLSCHAFT
FÜR INFORMATIK

Klaus David, Kurt Geihs, Martin Lange, Gerd Stumme (Hrsg.)

INFORMATIK 2019

50 Jahre Gesellschaft für Informatik
–
Informatik für Gesellschaft

Konferenzbeiträge der 49. Jahrestagung der Gesellschaft für Informatik

23.-26.9.2019
Kassel, Deutschland

Gesellschaft für Informatik e.V. (GI)

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

Volume P-294

ISBN 978-3-88579-688-6
ISSN 1617-5468

Volume Editors

Prof. Dr. Klaus David
Universität Kassel
Fachgebiet Kommunikationstechnik
Wilhelmshöher Allee 71-73
34121 Kassel
Email: david@uni-kassel.de

Prof. Dr. Kurt Geihs
Universität Kassel
Fachgebiet Verteilte Systeme
Wilhelmshöher Allee 71-73
34121 Kassel
Email: geihs@uni-kassel.de

Prof. Dr. Martin Lange
Universität Kassel
Fachgebiet Theoretische Informatik / Formale Methoden
Wilhelmshöher Allee 71-73
34121 Kassel
Email: mlange@uni-kassel.de

Prof. Dr. Gerd Stumme
Universität Kassel
Fachgebiet Wissensverarbeitung
Wilhelmshöher Allee 71-73
34121 Kassel
Email: stumme@uni-kassel.de

printed by Köllen Druck+Verlag GmbH, Bonn

Vorwort

„50 Jahre Gesellschaft für Informatik – Informatik für Gesellschaft“ lautete das Leitthema der INFORMATIK 2019, der 49. Jahrestagung der Gesellschaft für Informatik (GI). Dieses Leitthema drückt aus, dass die im Jahr 1969 gegründete GI in 2019 ihr 50. Gründungsjubiläum feierte, und es reflektiert, dass der gesellschaftliche Bezug der Informatik sowohl ein zentrales Anliegen der GI als auch der Informatik an der Universität Kassel darstellt. Das an der Universität Kassel angesiedelte interdisziplinäre Wissenschaftliche Zentrum für Informationstechnikgestaltung (ITeG) war daher maßgeblich an der Organisation der Tagung beteiligt. Das Tagungsmotto passt auch zum allgemein interdisziplinär geprägten Leitbild der Universität Kassel in einer aufstrebenden Region, die sich durch ein produktives Miteinander von Kultur, Wissenschaft und Wirtschaft auszeichnet.

Gegenüber früheren GI-Jahrestagungen erlebte die INFORMATIK 2019 in Kassel die Instanziierung eines neuen Tagungsformats. Kern des Programms waren sieben wissenschaftliche Tracks zu aktuellen Themenfeldern, denen besondere Aufmerksamkeit in Wissenschaft, Praxis und Gesellschaft zukommt. Gleichzeitig wurde die Zahl der Workshops deutlich reduziert. Die Organisation der Begutachtung und Auswahl der Beiträge in den Tracks lag in Händen von jeweils zwei Track Chairs. Vier eingeladene Keynotes und eine Präsentation zur Kunst (in) der Informatik rundeten als Plenumsveranstaltungen das Vortragsprogramm ab.

Als Ergebnis des Aufrufs zur Einreichung von Beiträgen lagen den Programmkomitees der Tracks insgesamt 122 Beiträge vor. Zu jedem Beitrag wurden mehrere Gutachten erstellt. Als Ergebnis wurden insgesamt 71 Beiträge (43 Vollbeiträge und 28 Kurzbeiträge), d.h. 58%, zur Präsentation auf der Tagung angenommen. Die Tagungsveranstalter waren sehr positiv überrascht vom Echo auf den neuen Tagungsschwerpunkt und das neue Tagungsformat, das als wegweisend für zukünftige Jahrestagungen gelten kann. Der vorliegende Tagungsband enthält die angenommenen Beiträge sowie die erweiterten Zusammenfassungen der eingeladenen Vorträge.

Der besondere Dank der Herausgeber dieses Tagungsbandes gilt den Track Chairs für die Organisation der Begutachtung und die Auswahl der Beiträge, den Mitgliedern der Programmkomitees und – last but not least - den Autoren der Beiträge. Ebenfalls gebührt unser Dank den Sponsoren für die großzügige Unterstützung der Veranstaltung sowie den vielen hier nicht genannten Helferinnen und Helfern für ihren Einsatz bei der Vorbereitung und Durchführung der Tagung.

Kassel, im September 2019
Klaus David, Kurt Geihs, Martin Lange, Gerd Stumme
(die Herausgeber)

Sponsoren

Wir danken den folgenden Unternehmen und Institutionen für die Unterstützung der INFORMATIK 2019.

Micromata GmbH Kassel

OctaVIA AG, Kassel

FLAVIA IT Management GmbH, Kassel

Yatta Solutions GmbH, Kassel

SMA Solar Technology AG, Niestetal

Deutsche Telekom IT-GmbH, Bonn

Hübner GmbH & Co. KG, Kassel

DE GRUYTER, Oldenburg

Springer Vieweg, Wiesbaden

plentysystems AG, Kassel

IBM Deutschland Research & Development GmbH, Böblingen

genua GmbH, Kirchheim bei München

Universität Kassel

Wissenschaftliches Zentrum ITeG an der Universität Kassel

Tagungsleitung

Gesamtleitung:	Prof. Dr. Kurt Geihs, Universität Kassel (Sprecher)
	Prof. Dr. Klaus David, Universität Kassel
	Prof. Dr. Martin Lange, Universität Kassel
	Prof. Dr. Gerd Stumme, Universität Kassel
Workshops:	Prof. Dr. Claude Draude, Universität Kassel
	Prof. Dr. Bernhard Sick, Universität Kassel
Firmendialog:	Prof. Dr. Albert Zündorf, Universität Kassel

Wissenschaftliche Leitung

Track 1 - Socio-technical Design and Value Orientation	Prof. Dr. Walid Maleej, Universität Hamburg Prof. Dr. Alexander Felfernig, TU Graz
Track 2 – Internet of Everything	Prof. Dr. Anna Förster, Universität Bremen Prof. Dr. Matthias Wählisch, FU Berlin
Track 3 – Data Science	Prof. Dr. Ingo Scholtes, Universität Zürich Prof. Dr. Markus Strohmeier, RWTH Aachen
Track 4 - Informatik mit Recht	Prof. Dr. Rüdiger Grimm, Universität Koblenz-Landau Prof. Dr. Gerrit Hornung, Universität Kassel Prof. Dr. Christoph Sorge, Universität des Saarlandes Prof. Dr. Indra Spieker gen. Döhmann, Univ. Frankfurt Prof. Dr. Arno Wacker, UniBW München
Track 5 – Sicherheit, Zuverlässigkeit, Korrektheit	Dr. Juliane Krämer, TU Darmstadt Prof. Dr. Roland Meyer, TU Braunschweig
Track 6 - Digitalisierung des Energiesystems	Prof. Dr. Christoph Krauß, Fraunhofer SIT, Darmstadt Prof. Dr. Kurt Rohrig, Fraunhofer IEE, Kassel
Track 7 – Digitale Bildung	Prof. Dr. Nadine Bergner, TU Dresden Prof. Dr. Ira Diethelm, Universität Oldenburg

Inhaltsverzeichnis

Keynotes

Track 1 - Socio-technical Design and Value Orientation

Full Papers

Extended Abstracts

Track 2 – Internet of Everything

Full Papers

Extended Abstracts

Track 3 – Data Science

Full Papers

Extended Abstracts

Track 4 - Informatik mit Recht

Full Papers

Track 5 – Sicherheit, Zuverlässigkeit, Korrektheit

Full Papers

Track 6 - Digitalisierung des Energiesystems

Invited Presentations

Full Papers

Track 7 – Digitale Bildung

Full Papers

Anhang

Keynotes

Learning causal mechanisms

Bernhard Schölkopf[1]

Abstract: In machine learning, we use data to automatically find dependences in the world, with the goal of predicting future observations. Most machine learning methods build on statistics, but one can also try to go beyond this, assaying causal structures underlying statistical dependences. Can such causal knowledge help prediction in machine learning tasks? We argue that this is indeed the case, due to the fact that causal models are more robust to changes that occur in real world datasets. We discuss implications of causal models for machine learning tasks, focusing on an assumption of 'independent mechanisms', and discuss an application in the field of exoplanet discovery.

[1] Max-Planck-Institut für Intelligente Systeme, Tübingen bs@tuebingen.mpg.de

doi:10.18420/inf2019_01

Smart Cities – Making Cities Livable and Sustainable

Siobhán Clarke[1]

Abstract: Given growing urban populations, it is clear we need to change our behaviour to better manage the sharing of increasingly constrained urban resources, such as the road network, energy, water, and so on. With an expected 70% of the world's population living in urban areas by 2050, pressure on resources and infrastructure in cities and communities around the globe is growing. Cities consume over two-thirds of the world's energy and account for more than 70% of global CO2 emissions. In an analysis of 13,000 cities published in 2018, the critical impact city dwellers have on overall carbon emissions is clear, and even more interestingly, it could be argued that city planning is hugely influential as it was found that roughly one third of an urban resident's footprint is determined by that city's public transportation options and building infrastructure [Mo18]. Pressure on city resources is clearly affecting quality of life, adversely impacting the environment and limiting economic growth.

Significant advances have been made in recent years relating to high-bandwidth network connectivity and highly-instrumented cities providing real-time information about the state of a city's resources. These technologies can be exploited to enable cities to work better. This talk explores how automation, using real-time decision-making, can play a part in assisting citizens in making better use of the resources available to them. The goal is not to take over citizens' lives, but to remove the onus on citizens to be constantly aware of potential opportunities for optimising resource sharing. In particular, the talk draws on our recent research, using examples from autonomous vehicles [MBL19], vehicle sharing [GC18] and energy demand-side management [Ma19].

References

[GC18] Golpayegani, Fatemeh; Clarke, Siobhán: Co-Ride: Collaborative Preference-Based Taxi-Sharing and Taxi-Dispatch. In (Tsoukalas, Lefteri H.; Grégoire, Éric; Alamaniotis, Miltiadis, eds): IEEE 30th International Conference on Tools with Artificial Intelligence, ICTAI 2018, 5-7 November 2018, Volos, Greece. IEEE, pp. 864–871, 2018.

[Ma19] Marinescu, Andrei; Taylor, Adam; Clarke, Siobhán; Serban, Ioan; Marinescu, Corneliu: Optimising residential electric vehicle charging under renewable energy: Multi-agent learning in software simulation and hardware-in-the-loop evaluation. International Journal of Energy Research, 43(8):3853–3868, May 2019.

[MBL19] Monteil, J.; Bouroche, M.; Leith, D. J.: $\mathcal{L}_2$ and $\mathcal{L}_\infty$ Stability Analysis of Heterogeneous Traffic With Application to Parameter Optimization for the Control of Automated Vehicles. IEEE Transactions on Control Systems Technology, 27(3):934–949, May 2019.

[1] Trinity College Dublin, Ireland siobhan.clarke@scss.tcd.ie

doi:10.18420/inf2019_02

[Mo18] Moran, Daniel; Kanemoto, Keiichiro; Jiborn, Magnus; Wood, Richard; Többen, Johannes; Seto, Karen C: Carbon footprints of 13 000 cities. Environmental Research Letters, 13(6):064041, jun 2018.

GI 50 – und wie geht es weiter?

Volker Claus[1] Stefan Jähnichen[2] Reinhard Wilhelm[3]

Abstract: Mit der Aussage „Alles fließt" formulierte Heraklit vor 2500 Jahren, dass sich alles verändert - damals recht langsam, heute jedoch sehr schnell. Als vor 50 Jahren die Gesellschaft für Informatik gegründet wurde, stellte sich die Informatik als eine Grundlagenwissenschaft dar, die vor allem den Arbeitsbereich erleichterte und erweiterte. Mit dem Vordringen in immer neue Anwendungsbereiche und dem explosionsartigen Anwachsen der IT-Industrie traten die ingenieurwissenschaftlichen Aspekte, neue Kommunikationsformen und Veränderungen des Verhaltens, Verwaltens und Gestaltens in den Vordergrund und verlangen nach einer GI, die diese Weiterentwicklung vorantreibt, formt und kritisch begleitet.

Dafür braucht die GI eine möglichst große Anzahl an Mitgliedern. Doch wie viele Organisationen, Parteien, Verbände verzeichnet auch die GI einen Mitgliederschwund. Vor allem die Jugend organisiert sich lieber über soziale Medien und schätzt es, sich ohne lokale, regionale oder nationale Begrenzungen mit Menschen ähnlicher Interessenslagen zu vernetzen. Dabei werden entstehende Risiken und Gefahren oftmals ignoriert. Wie soll sich die GI hierauf einstellen und ihre Attraktivität in Zukunft steigern? Oder soll sie sich auf ihr „Kerngeschäft", die Wissenschaft und ihre Umsetzungen beschränken und insbesondere die fachlichen Kompetenzen ihrer Mitglieder stärken?

Anders gefragt: Wenn wir heute eine neue GI gründen würden, welche Aufgaben, Konzepte und Werte würden wir ihr mitgeben und welche Struktur und Regeln sollte sie haben, um ihre Ziele zu erreichen, welche Wege sollte sie beschreiten, um dem Menschheitswohl zu dienen, um ihre Mitglieder zu fördern, um kompetent, sichtbar, vernetzt, einflussreich, verantwortlich, demokratisch, professionell, einflussreich usw. zu sein? Wie betten sich die bisherigen 50 Jahre an Leistungen, Erfolgen und Erfahrungen der GI in solche Visionen ein? Der Vortrag regt Wege in verschiedene Richtungen an, wie die nächsten 50 Jahre in Angriff genommen werden können (oder zumindest die nächsten 10 Jahre, denn die Veränderungsgeschwindigkeit scheint ungebremst zu wachsen).

[1] Universität Stuttgart, Fakultät 5, Fachbereich Informatik, volker.claus@informatik.uni-stuttgart.de
[2] Technische Universität Berlin, Fakultät IV - Elektrotechnik und Informatik, stefan.jaehnichen@tu-berlin.de
[3] Universität des Saarlandes, Saarland Informatics Campus, wilhelm@cs.uni-saarland.de

doi:10.18420/inf2019_03

Data Science for Social Good: Opportunities and Challenges

Ciro Catutto[1]

Abstract: The value of big data and advanced analytics lies critically in the opportunity to make better decisions and to design better policies. Identifying needs, targeting interventions, and measuring impact are all challenges that can greatly benefit from more quantitative approaches and data-intensive methods. This opportunity is currently stimulating new research lines in academia, new data sharing initiatives in industry, and new programs in the non-profit sector, while also calling for novel cross-sector collaborations around data. This talk will reflect on the complex interplay of new data sources, data science methods and algorithmic decisions, discussing selected case studies in the domains of health and mobility, and highlighting opportunities as well as challenges for the generation of public value and social impact.

[1] ISI Foundation ciro.cattuto@isi.it

doi:10.18420/inf2019_04

Computer, Kunst und Künstlichkeit

Frieder Nake[1]

Abstract: So weit zurück ich auch blicke, immer wieder tauchte sie auf, die merkwürdige Frage: "Wer ist es denn nun, der die Kunst macht?" Schon vor Gründung der GI wurde so gefragt. Und immer dann tritt die ziemlich gleichbleibende Frage mit erneuter Dringlichkeit wieder auf, wenn ein besonderes Ereignis die Aufmerksamkeit des Publikums auf sich zieht. So, wie vor kurzem der spektakulär hohe Verkaufs-Preis von $432,000 für ein Bild, das aus dem Computer kam. Die Frage taucht auf, sobald in Galerien Zeichnungen ausgestellt werden, bei deren Produktionsprozess Algorithmen, Programme oder ausführende Computer auf neue Weise verwendet werden. Das geschieht seit mehr als 50 Jahren und derzeit geschieht es unter den Flaggen von Big Data und Neuronalen Netzen. – Die Frage hat, so scheint mir, eine immer gleich bleibende Antwort, auch wenn deren Kontext sich drastisch ändern mag. Die Antwort ist beharrlich, dass der Computer es nicht ist, der die Kunst macht, was kaum überraschen wird. Im Vortrag soll dies an Beispielen erörtert und in einen kultur-kritischen Zusammenhang gestellt werden. Was nämlich können wir aus dem ästhetischen und kunsthistorischen Diskurs lernen über Merkwürdigkeiten der spekulativen Zuschreibungen, die die Operationen von Computern und Hervorbringungen der Informatik zu umwehen scheinen? Computer werde ich als *semiotische Maschinen*, Menschen als *semiotische Tiere* (Felix Hausdorff) kennzeichnen. Im gemeinsamen Bezug auf Zeichenprozesse ist der Grund zu suchen für das immer wieder aufflackernde Vergnügen an der Menschwerdung des Computers, der doch nichts als eine Maschine ist, von besonderer Art zwar, aber eben doch Maschine. Einen kleinen Rempler gegen die »Künstliche Intelligenz« werde ich mir nicht versagen können.

[1] Universität & Hochschule für Künste, Bremen, nake@uni-bremen.de

doi:10.18420/inf2019_05

Track 1 - Socio-technical Design and Value Orientation

Socio-Technical Design and Value Orientation

Walid Maalej[1], Alexander Felfernig[2]

"Software systems are developed by humans for humans". This motto represents the main driver of this track. On one hand, social and human factors influence Software Engineering activities (and their productivity) as well as software systems (and their quality). Particularly empirical research has aimed during the last decade to understand, leverage, and consider human and social factors when developers, testers, managers, and users interact in software projects. On the other hand, software is pervasive in our lives: it mediates people-to-people communication, supports human choices, and might even have far-reaching impact on lives, economies, and the planet. Software and its development needs to accommodate a wide range of social and human values, such as trust, governance, reputation, privacy, and sustainability – which by itself should be reflected in design, engineering, and deployment processes.

This track brings together the core contributing communities on socio-technical design and value orientation to present and discuss cutting edge research and to further advance the field. We particularly target the communities of Requirements Engineering, Software Engineering, Information Systems, CSCW/Social Computing, and Societal Computing but are also seeking to cross boundaries to related fields.

Topics of interest for this track included: human aspects in software Engineering; socio-technical design; requirements engineering, in particular value-driven RE; theories and applications of social and crowd computing; sustainability, explainability and trust of complex software systems; engineering social systems; software engineering impact on society; value of software systems and processes; ethical and legal aspects in software development; user involvement; feedback and interaction loops: between systems, users, and developers.

The program committee correspondently consisted of experts in the related area (with focus on German speaking experts as the conference does):

- Eva Bittner, University of Hamburg
- Claude Draude, University of Kassel
- Jörg Dörr, Fraunhofer IESE

[1] University of Hamburg, maalej@informatik.uni-hamburg.de
[2] TU Graz, alexander.felfernig@ist.TUGraz.at

doi:10.18420/inf2019_06

- Schahram Dustdar, TU Wien
- Gregor Engels, University of Paderborn
- Alexander Felfernig, TU Graz (co-chair)
- Martin Glinz, University of Zurich
- Eric Knauss, Chalmers and University of Gothenburg
- Agnes Koschmider, KIT Karlsruhe
- Walid Maleej, University of Hamburg (co-chair)
- Daniel Mendez, TU Munich
- Judith Michael, RWTH Aachen
- Ali Sunyaev, KIT Karlsruhe
- Barbara Paech, University of Heidelberg
- Birgit Penzenstadler, California State University
- Guenther Ruhe, University of Calgary
- Christoph Stanik, HITeC e.V.

The track solicited two kinds of submissions. First, extended abstracts provide a summary of already published outstanding work, which recently appeared in top international venues and which should be presented to the German-speaking community. Second, regular articles present original research with new insights and stable results. In total, we received 5 abstracts published at top venues (ACM Transaction on Social Computing, ACM Conference on Computer Supported Cooperative Work (CSCW), International Conference on Software Engineering (ICSE), and International IEEE Conference on Requirements Engineering (RE)). We accepted all five abstracts for presentations at the conference. For the details, the readers of this proceedings are redirected in the abstracts to the original publications. Moreover, we received 12 full paper submissions. One was incomplete and thus desk-rejected. Each of the remaining 11 submissions had three detailed reviews by different members of the program committee. At the end we accepted 4 full papers based on the reviews and a follow-up discussion.

We hope that this track – including the presentations of accepted papers, the invited talks, and the informal interactions and discussions – will contribute to building a "new" community around this central, emerging, multi-disciplinary topic and that we will see follow up projects and venues in future.

Full Papers

Ethical decomposition as a new method to analyse moral dilemmata

Findings on mad trolleys and self-driving cars

Christian Thielscher[1], Bianca Krol[2], Stefan Heinemann[3], Michael Schlander[4]

Abstract
Introduction: Since P. Foot studied the trolley problem in 1967, it has been extensively discussed in ethics, decision theory, medicine, and other disciplines. With the invention of autonomous vehicles, it has become an important and urgent practical question.
Methods: Three well-known, one new and one slightly changed versions of the trolley problem are arranged in a specific order. Kantian and utilitarian solutions to the problems are discussed. Respondents' decisions were empirically tested and aligned with ethical theories.
Results: Both Kantian and utilitarian ethics provide rules for decision. However, both are incomplete and differ from each other sometimes. In one case, both recommend to "act", in another, "not to act". In these two cases, almost all respondents follow the mutual advice. In other cases, ethical theories as well as responses differ.
Discussion: Respondents did not behave irrationally; rather, they considered ethical theories in a sensible way. Disaggregating Kantian and utilitarian decisions helps to identify situations in which autonomous vehicles can be programmed for better adaption to user preferences.

Keywords: Self-Driving Cars, Moral Dilemmata, Kantian Ethics, Utilitarian Ethics, Experimental Ethical Research.

1 Introduction

Since P. Foot studied the trolley problem in 1967 [Fo67], it has been extensively discussed in ethics, decision theory, medicine, and other disciplines [Th17]. In particular, the question of why people decide differently in the trolley case vs. the fat man case (see below) has received a lot of attention [Ed15]. With the invention of autonomous vehicles – especially level 5 (SAE International's standard J3016) autonomous cars – it has become an important and urgent practical question [BSR16].

In Germany, for example, the "Ethics Commission on Automated Driving" recently recommended (amongst others) [BM18]:

[1] FOM University of Applied Sciences, Competence Center for Medical Economics, Leimkugelstr. 6, D-45141 Essen, christian.thielscher@fom.de
[2] FOM University of Applied Sciences, Institute for Empirical Research & Statistics, Leimkugelstr. 6, D-45141 Essen, bianca.krol@fom.de
[3] FOM University of Applied Sciences, Leimkugelstr. 6, D-45141 Essen/University Medicine Essen, stefan-heinemann@gmx.de
[4] German Cancer Research Center (DKFZ), Department of Health Economics, Im Neuenheimer Feld 581, D-69120 Heidelberg, michael.schlander@dkfz.de

doi:10.18420/inf2019_07

- Automated and connected driving is an ethical imperative if the systems cause fewer accidents than human drivers (positive balance of risk).
- Damage to property must take precedence over personal injury. In hazardous situations, the protection of human life must always have top priority.
- In the event of unavoidable accident situations, any distinction between individuals based on personal features (age, gender, physical or mental constitution) is impermissible.

The trolley problem addresses a thought experiment where a decision-maker is faced with an ethical dilemma. Which is the morally correct choice in a scenario where a runaway trolley on railway tracks is headed for five people who are unable to move:

1. Do nothing and the trolley kills five people?

2. Pull a lever and divert the trolley onto another track where it will kill only one person?

There are many versions of this "sacrifice-one-to-rescue-five"-dilemma including various narratives and numbers of options. The forerunner of the autonomous trolley problem was J. J. Thomsons' paper from 1985 [Th85], whereas in Foots' original version from 1967, the trolley was manned.[5] Foots' text reads as follows [Fo67]:

"Suppose that a judge or magistrate is faced with rioters demanding that a culprit be found for a certain crime and threatening otherwise to take their own bloody revenge on a particular section of the community. The real culprit being unknown, the judge sees himself as able to prevent the bloodshed only by framing some innocent person and having him executed. Beside this example is placed another in which a pilot whose aeroplane is about to crash is deciding whether to steer from a more to a less inhabited area. To make the parallel as close as possible it may rather be supposed that he is the driver of a runaway tram which he can only steer from one narrow track on to another; five men are working on one track and one man on the other; anyone on the track he enters is bound to be killed. In the case of the riots the mob has five hostages, so that in both the exchange is supposed to be one man's life for the lives of five. The question is why we should say, without hesitation, that the driver should steer for the less occupied track, while most of us would be appalled at the idea that the innocent man could be framed."

P. Foot also invented the problem which was later called "transplant":

"Why…do we not feel justified in killing people to obtain…spare parts for grafting on to those who need them?"

This question is at the heart of the trolley problem. P. Foot proposes that there are "positive" and "negative" duties. She writes [Fo67]:

[5] There have been earlier versions, e g. [We51].

"The steering driver faces a conflict of negative duties, since it is his duty to avoid injuring five men and also his duty to avoid injuring one. In the circumstances, he is not able to avoid both, and it seems clear that he should do the least injury he can. The judge, however, is weighing the duty of not inflicting injury against the duty of bringing aid. He wants to rescue the innocent people threatened with death but can do so only by inflicting injury himself. Since one does not in general have the same duty to help people as to refrain from injuring them, it is not possible to argue to a conclusion about what he should do from the steering driver case. It is interesting that, even where the strictest duty of positive aid exists, this still does not weigh as if a negative duty were involved. It is not, for instance, permissible to commit a murder to bring one's starving children food. If the choice is between inflicting injury on one or many there seems only one rational course of action; if the choice is between aid to some at the cost of injury to others, and refusing to inflict the injury to bring the aid, the whole matter is open to dispute. So, it is not inconsistent of us to think that the driver must steer for the road on which only one man stands while the judge (or his equivalent) may not kill the innocent person in order to stop the riots."

According to this logic, we may not kill an uninvolved person in order to rescue five others with his organs (as in the "transplant" case); however, the pilot and the trolley driver may kill one because otherwise they would kill five – here two negative duties collide and in this case, we kill as few as possible.

Thomson however, objects to Foots solution; she disagrees with the sentence "killing one is worse than letting five die". In order to prove its falsity, she introduced both the "bystander at the switch" as well the "fat man problem" (Foot had a different fat man problem in her paper) [Th85, p. 1397]:

"Let us begin by looking at a case that is in some ways like Mrs. Foot's story of the trolley driver. I will call her case Trolley Driver; let us now consider a case I will call Bystander at the Switch. In that case, you have been strolling by the trolley track, and you can see the situation at a glance: The driver saw the five on the track ahead, he stamped on the brakes, the brakes failed, so he fainted. What to do? Well, here is the switch, which you can throw, thereby turning the trolley yourself. Of course, you will kill one if you do. But I should think you may turn it all the same."

(The authors of this paper disagree with Thompson on the bystanders' recommended behavior as we will explain in more detail later.)

Thompson also introduced the "fat man" problem, and the problem which it creates [Th85, p. 1409].

"Consider a case - which I shall call 'Fat Man' - in which you are standing on a footbridge over the trolley track. You can see a trolley hurtling down the track, out of control. You turn around to see where the trolley is headed, and there are five workmen on the track where it exits from under the footbridge. What to do? Being an expert on trolleys, you know of one certain way to stop an out-of-control trolley: Drop a really heavy

weight in its path. But where to find one? It just so happens that standing next to you on the footbridge is a fat man, a really fat man. He is leaning over the railing, watching the trolley; all you have to do is to give him a little shove, and over the railing he will go, onto the track in the path of the trolley. Would it be permissible for you to do this? Everybody to whom I have put this case says it would not be. But why?"

Again, the key question is: why would most people throw the switch but not shove the fat man? And should an autonomous vehicle in a comparable situation – for example if it has to decide to either kill five children playing on a street or run into a pedestrian and kill him – decide like in a "fat-man-problem" or in the trolley case?

Thomson gives several potential explanations. First, she says that shoving is a direct infringement of the fat man's rights whereas throwing the switch isn't. However, as she herself notices, this explanation does not work. Imagine that you do not have to shove the fat man but just need to wobble the handrail: you still would not – although "wobbling" the handrail and "throwing" the switch are pretty close (at least morally). Thomson finally explains [Th85, p. 1410]:

"So the means by which the agent in 'Fat Man' gets the trolley to threaten one instead of five include toppling the fat man off the footbridge; and doing that is itself an infringement of a right of the fat man's. By contrast, the means by which the agent in 'Bystander at the Switch' gets the trolley to threaten one instead of five include no more than getting the trolley off the straight track onto the right-hand track; and doing that is not itself an infringement of a right of anybody's."

The authors of this paper do not accept this explanation. If one throws the switch he definitely kills the one person on the other track; it is hard to see how this is "no infringement" of his right to live. We also do not accept other solutions that were proposed, e.g., that wobbling the handrail requires the decider to be close to the victim of his doing (the fat man); whereas the switch thrower is distant from the killed person. It is easy to construct a specific case where a fat man stands on a trap door and a trigger can be pulled from somewhere to open it (however, still in this case, most people would not kill the fat man but would throw the switch). Finally, we do not think that people behaving differently in these situations are simply inconsistent in their behavior.

We believe that both P. Foot as well as Thomson point into the right direction. Our key hypothesis is the idea that "negative duties overrule positive duties" is a Kantian rule (see below) whereas the central utilitarian rule is "kill as few as possible". Respondents simply use both rules, if they agree, and, if they disagree, prefer one of these over the other, depending on the specific details of the story. If they agree, autonomous vehicles should as well.

By ethical decomposition of five test cases, we prove that

- Kantian and utilitarian decision rules agree that it is better to act – to throw the switch etc. – in one case (and, therefore, the majority of respondents does act),

- agree not to act in another case (so that respondents don't act), and
- disagree in the other cases (so that some respondents act while others don't). We finally test the hypothesis empirically.

2 Methods

In light of the various possibilities and potential issues and solutions relating to the trolley narrative, this paper aims to explore the decision of people in different variations of the trolley case. To do so we devised a set of cases - by developing one new case, changing another one and rearranging them together with three known cases - which serve to

- create a continuum of changing decisions so that almost all responders would "act" (throw the switch, shove the fat man, etc.) in the first case and almost none would act in the last case; and
- differentiate between Kantian and utilitarian perspectives. We used these two because they are seen as the most influential non-religious ethical theories [Sa10]. Of course, ethical decomposition works with other theories as well.

Finally, we empirically tested the cases with respondents. Responses were analyzed with chi square test.

3 Results

There are five cases in our continuum:

1.: In the "dynamite wagon" case the decision-maker is standing near the trolley track seeing an unmanned trolley hurtling down the track, out of control. There are three tracks – one with one person chained to the track, one with five persons chained, and one with the dynamite wagon. What to do? If the decision-maker does nothing, the unmanned trolley will run into the dynamite wagon loaded with explosives and the detonation will kill both the one person on one track, as well as the group of five people on the other track. However, the decision-maker can choose to throw the switch so that the trolley cart is diverted from the dynamite - but then kills either one or five.

2. and 3.: In addition to the "dynamite wagon", we propose two different versions of the trolley case: in the first one, the switch is loose so that it can move from one side to the other; if the trolley arrives it will cause the switch to point in one or the other direction with a 50/50 chance (this is what we call the „trolley, floppy switch" case). The traditional case where the switch points to the five is called "trolley, fixed switch".

4. and 5.: Finally, we use the "fat man" and "transplant" cases as described above.

The following table presents an overview of cases (table 1 – we will explain the column "expected response" in a second):

Problem	**Short description**	**Utilitarian decision**	**Kantian decision**	**Expected response**
1. Dynamite wagon	Trolley runs into dynamite wagon, thus killing all; bystander may throw switch in other direction, thus killing one or five	Throw switch	Throw switch	> 95%
2. Trolley, floppy switch	A trolley will kill either one or five; unclear position of switch; bystander may throw switch in either direction, thus killing one or five	Throw switch	Throw switch (?)	> 50%
3.Trolley, fixed switch	A trolley will kill five because switch points in their direction; bystander may throw switch in other direction, thus killing one	Throw switch	Do nothing	~ 50%
4. Fat man	A trolley will kill five unless you shove a fat man over bridge	Shove fat man	Do nothing	< 50%
5. Transplant	Five persons will die from disease unless you kill another person, transplanting his organs	Do nothing (?)	Do nothing	~ 0%

Tab. 1: Description of cases

We now discuss these cases in order to differentiate between the utilitarian and Kantian decision-making approach. By the "utilitarian" approach we mean Benthams' basic version where the sum of happiness is maximized [Be70]. With "Kantian" we refer to Kants' famous categorical imperative. Kant used different formulations which he thought are interchangeable. We employ two of these phrases [Ka61]:

1. “Handle nur nach derjenigen Maxime, durch die du zugleich wollen kannst, dass sie ein allgemeines Gesetz werde.” This can be loosely translated as: act on principles that are well-suited as foundation for a law.

2. “Handle so, daß du die Menschheit sowohl in deiner Person, als in der Person eines jeden andern jederzeit zugleich als Zweck, niemals bloß als Mittel brauchest.” That is, behave at all times such that you treat humanity both in the form of your person, as well as all others, as an end, but never merely as a means.

From the Kantian perspective, it is forbidden to kill an uninvolved person in order to save five because this would treat the individual as a means rather than as an end. This interpretation is close to Foots’ sentence: “killing one is worse than letting five die.”

In “dynamite wagon” both utilitarians, as well as Kantians will throw the switch. However, there is a problem for Kantians: in which direction? While it seems natural to kill one rather than five, it is not easy to see the principle which would allow for a law. For example, the principle “it is better to kill fewer people than more” cannot always be correct. Imagine you have to either rescue two Texan criminals sentenced to death and waiting for their penalty or one seven-year old girl.

“Everything else being equal if you have to kill either one or five then choose the one” seems to be a good candidate but still is far from being perfect. Imagine, for example, that a sadistic police officer in a totalitarian country forces you to decide: either you kill one of five innocent prisoners or he kills all five. Are you entitled to kill one? Which one? It seems as if the Kantian rule is indecisive in these cases – which is not very surprising since Kant did not want to develop a metric to decide in every dilemma; rather he wanted to identify ethical rules that are always correct (independent of the situation).

In “floppy switch”, utilitarians will clearly throw the switch as well, whereas Kantian logic is again open for discussion. One could argue that a Kantian might be hesitant to throw the switch, being afraid of using the one person as a means only. However, if he does not, he lets coincidence decide; in this case he uses the one or the five killed as fortuities’ means to save the other(s). Before the floppy switch becomes fixed (by chance), both the one as well as the five persons are “involved” by being threatened. If the Kantian throws the switch, he has to treat either the one or the five as a means and the other(s) as an end. In this conflict, some Kantians will feel entitled to throw the switch while others do not. It is close to the situation where a pilot has to decide where to crash his broken plane: in an inhabited area or redirect it into a less peopled region. Kantians willing to change the planes’ direction will point at the fact that people living close to aeroplane routes know that they might get into the way of busted machines, that pilots will try to kill as few as possible; and therefore, implicitly accept the pilots’ behavior.

In “trolley, fixed switch”, Kantians would NOT throw the switch whereas utilitarians would. In this case, the one person is uninvolved (not threatened by coincidence) and therefore, Kantian logic would forbid to kill him.

This also makes clear that the ethical watershed runs between “floppy switch” and “fixed switch” (and not between “fixed switch” and “fat man”).

The authors think that this is an important finding. We believe that some listeners who prefer Kantian ethics may mix “trolley, floppy switch” with “trolley, fixed switch” and treat the latter as if it was the former: it is easy to confound the two. Kantians who think about “fixed switch” may misunderstand that the switch indeed is fixed; in this case, they may think they are in “floppy switch” and decide to throw the switch – although they should not. By clarifying the situation, Kantians who threw the switch in “fixed switch” erroneously (because they thought it was a “floppy case”) will change their mind. Our empirical results point into this direction.

In fat man, pure utilitarians would shove or wobble, and even in transplant it is difficult for them to find a way not to kill the one person: they could, e.g., say that if they did people might not enter hospitals anymore (because they expect to get killed and eviscerated there), thus causing trouble and loss of happiness for the society. However, this introduction of indirect societal costs and benefits (that is, costs that are not directly linked to the case as such but rather potential consequences of it) creates significant uncertainties: which indirect costs are to consider and how are they valued? Where to start and stop with calculation of potential consequences? E.g., would people use bridges (like in the “fat man problem”) if they think they get shoved?

Therefore, neither Kantian nor utilitarian rules provide clear decisions on all of the cases. In addition, sometimes they agree whereas sometimes they disagree depending on the respective variation of the basic conflict (kill one to rescue five).

We empirically tested this intuition with responses to our cases. In July 2018, we discussed the cases from “dynamite wagon” to “transplant” with 50 students. All of them studied medicine at the university hospital of Heidelberg, all being in their 5th semester. First, we introduced the cases to them and asked them whether they would “act” (i.e., depending on the specific case, throw the switch, shove the fat man, or kill the transplant patient).

More specifically, we explained the “trolley, fixed switch” as follows: “An unmanned, out-of-control trolley runs towards five people who are chained to the track; it will kill all five. You are standing close to a switch. If you throw it, the trolley changes direction and runs into one person who is chained to the track, killing this one but saving the other five.” We also showed a little drawing (Fig. 1). We then asked the students whether they were familiar with the case or not.

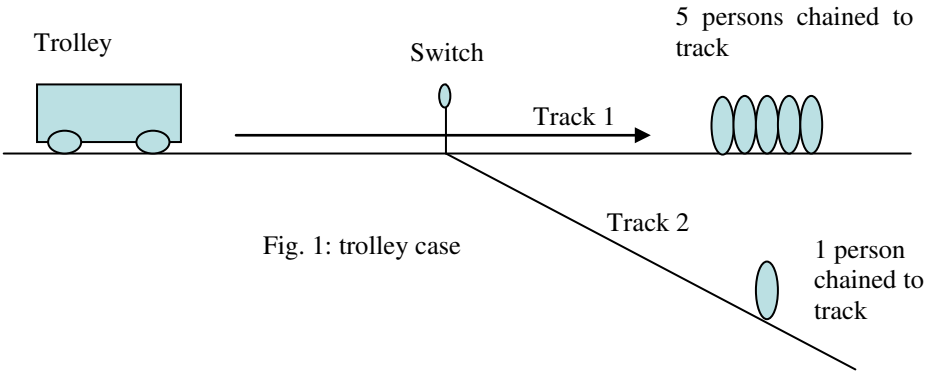


Fig. 1: trolley case

We continued by presenting the other cases (from "dynamite wagon" to "transplant").

Following this, the two groups of students (familiar / new to the problem) were asked to vote whether they would act (throw the switch, shove the fat man, etc.).

Afterwards, utilitarian vs. Kantian ethics and their perspective on the cases were discussed.

Finally, we had the same votes as before.

These are our findings (table 2):

	New to problem (n=30)		Familiar with problem (n=20)	
	before discussion	after discussion	before discussion	after discussion
Dynamite wagon***	21:3	23:1	16:0	16:0
Trolley, floppy switch***	21:9	23:4	12:6	12:6
Trolley, fixed switch***	23:5***	9:9***	17:2***	11:9***
Fat man***	7:18	8:17	4:12	3:15
Transplant***	0:25	0:25	0:19	0:19

Tab. 2: Results

The numbers in the boxes denote the number of students who act versus the number of students who do not. For example, in the group of students who were new to the problem, before discussion 21 would throw the switch in "dynamite wagon" whereas three would not. The numbers do not always add to the total number of students because some did not decide.

Results are statistically highly significant at the 0,001 level:

- All lines differ significantly from all other lines (by comparing the sum per line) – with one exception: trolley, floppy switch vs. trolley, fixed switch. However, in the group that was new to the problem, after discussion, those two cases did differ significantly.
- For both groups, the discussion did not change the responses significantly – again with one exception: both groups changed their responses significantly in trolley, fixed switch.

4 Discussion

The purpose of our study was to better understand ethical intuition and empirical responses to dilemma situations and to deduct recommendations for autonomous vehicles.

The overall pattern of our students' response is in line with expectations. With both Kantians and utilitarian ethicists throwing the switch in "dynamite wagon" (we didn't ask in which direction), we expected students to do the same – and the vast majority did. In "transplant", Kantians and (most) utilitarians would not act, and therefore, most respondents do not act as well. We conclude that in these cases, automatic control systems should do the same; guidelines for software developers might help to align engineering with unambiguous moral decisions. "Rolling dice" seems not to be the best strategy in these cases.

In the cases from "trolley" to "fat man", however, students need to decide which of two conflicting ethical rules to follow. Therefore, it was expected that the voting is in between the two ends of our continuum.

A bit of a surprise is the fact that before discussion, less students did not act in „fixed switch" than in "floppy switch"; this may be due to some uncertainty. However, after discussion, this effect disappeared.

Another interesting finding is that after discussion, more students did not act in "fixed switch". It seems that the comparison of the cases and ethical background fosters Kantian behavior.

Until now, the fact that people decide differently in comparable cases as we modelled here (all being versions of "kill one to rescue five") is often seen as a case of inconsistent behavior. An often-cited example for irrationalities is N. Taleb who highlighted incon-

sistencies in the thought processes of subjects when forecasting / estimating the likelihood of various scenarios [Ta08]. He calls this "pathologies in decision-making", and he also talks about "blindness" and the "paradox of perception", as well as "the pull of the sensational" and "the use of narrative to get attention". Taleb maps and separates general attributes of thinking and reasoning into what he calls System 1 and System 2. System 1 is experiential. It is associated with intuition that produces short cuts called "heuristics" that allow rapid action. System 2 is the cognitive one that we normally call "thinking" that is effortful, reasoned, logical, self-aware, etc. – the stuff that apparently lies at the basis of research and goes on in the classroom. This idea is also at the core of D. Kahneman's work on fast and slow thinking [Ka11].

However, there is debate whether this is indeed a pathology. P. Railton recently argued that System 1, being based on prior experiences and having condensed them to simple rules of thumb is wrong only if for some reason the current situation does not fit to former learnings [Ra15].

Another case often cited as a typical example for "irrational" behavior is the fact that people tend to pay higher attention to known lives as opposed to unknown. If miners are trapped in Chile, much effort is provided to rescue them; although this money would have saved much more people if used for feeding the hungry. However, there might be a rational explanation for this as well; e.g., if respondents use two rules of thumb as the following: "try to help were you can" and "you cannot save everybody in the world so focus on what you can do (and know about)".

After ethical decomposition, the responses given by the students are not irrational but rather in line with ethical theories. Even if there is some irrational part to decisions (as is probably always the case) it is much less clear whether all of the changes (from act to don't act) are due to inconsistencies. There might be good reasons to act "Kantian" in one situation and "utilitarian" in another.

We therefore think that just leaving it open to the artificial intelligence engineers how their car behaves may be not the best solution [GM17]. At least it should be possible to decompose "Kantian" versus "utilitarian" decisions in many potential situations; and if they agree, engineers should agree as well. Of course, this will work for some but not all situations. For example, in situations involving humans versus properties the German Ethics Commission on Automated Driving demands (in a Kantian way) that damage to property must take precedence over personal injury [BM18]. However, this is not undisputed. R. A. Posner, for example, says it is better to kill a child than 100.000 sheep [Po79].

On a more practical note, we reason that in addition to those of utilitarians, the interests of Kantian deciders need to be reflected in creating unmanned cars as well: a utilitarian self-driving vehicle will kill as few as possible, whereas a Kantian car is simply not entitled to kill anybody (if avoidable). For example, it is acceptable to a utilitarian if sometimes the vehicle software crashes and kills people – given that the pleasure created by autonomous vehicles is higher than the loss resulting from the crashes and that the

costs of building the vehicle in a safe way are higher than the costs of the pleasure lost by accidents. It is, therefore, much more difficult to construct a Kantian car because it requires fail-safe IT, well-developed Artificial Intelligence systems and extensive testing and planning. (Kant would probably also note that voluntarily creating vehicles which have the capacity to kill people are not in line with ethical demands and that the engineers who constructed them might want to read his books.) Thus if we do not know whether at least some people prefer Kantian cars, we should not risk killing them by accidence.

Of course, trolley type cases are not the only problem software developers face; rather, there is a multitude of challenges, e.g., of technology, cost-benefit-assessments, etc. [Ro19]. The trolley case tells something about human perspectives on ethics, and if we understand them better, we can transfer them to automatic systems.

There are two key limitations to our results.

First our experimental design did not control for group effects (e.g. that some respondents wanted to decide in the same way as the majority did).

Second, we cannot match voting yet with Kantian or utilitarian perspective; that is, we do not know which of our students see themselves as Kantian or utilitarian (or else). We can only conclude this from their respective voting. We will design further studies to find out.

5 Conclusion

Moral dilemmata can be analyzed both from utilitarian as well as Kantian perspective. Sometimes they agree in their decisions, sometimes they disagree. In cases where they agree, respondents decide almost unambiguously, and automatic control systems should do the same. Here, “rolling dice” seems not to be the best strategy.

Our experiments support the hypothesis that humans use both utilitarian and Kantian decision criteria. In some cases, they prefer utilitarian over Kantian rules and vice versa. This may explain behavior that looks “irrational” at first glance.

Ethical decomposition of conflicts can help with the difficult undertaking to tell clear cases (where utilitarians and Kantians agree) from unclear ones. In addition, it supports analysis of our moral intuition. More research is needed to experimentally scrutinize moral behavior of humans, and to translate it into guidelines for software development.

Acknowledgements: The authors would like to thank Linda O’Riordan, Julia Richenhagen, and Piotr Zmuda for valuable input.

Bibliography

[Be70] Bentham, J.: An Introduction to the Principles of Morals and Legislation. Edited by Burns, J. H. and Hart, H. L. A. Oxford University Press, Oxford, 1970

[BM18] BMVI Press Release: Ethics Commission on Automated Driving presents report. https://www.bmvi.de/SharedDocs/EN/PressRelease/2017/084-ethic-commission-report-automated-driving.html, accessed: July 20th, 2018

[BSR16] Bonnefon, J. F.; Shariff, A.; Rahwan, I.: The social dilemma of autonomous vehicles. Science, 352(6293), 2016, p. 1573-1576.

[Ed15] Edmonds, D.: Would You Kill the Fat Man? The Trolley Problem and What Your Answer Tells Us about Right and Wrong. Princeton University Press, Princeton, 2015

[Fo67] Foot, P.: The Problem of Abortion and the Doctrine of the Double Effect. Oxford Review, no. 5, 1967, p. 5-15

[GM17] Gogoll, J.; Müller, J.: Autonomous Cars: In Favor of a Mandatory Ethics Setting. Sci Eng Ethics (2017) 23, p. 691-700

[Ka11] Kahneman, D.: Thinking, Fast and Slow. Farrar, New York, 2011

[Ka61] Kant, I.: Grundlegung zur Metaphysik der Sitten. Edited by Valentiner, T. Reclam Stuttgart, 1961

[Po79] Posner, R. A.: Utilitarianism, Economics, and Legal Theory. The Journal of Legal Studies, 1979 8:1, p. 132f.

[Sa10] Sandel, M. E.: Justice. Farrar, New York, 2010

[Ra15] Railton, P.: "Dual-Process" Models of the Mind and the "Statistical Victim Effect". In: Cohen, I . G. et al.: Identified versus Statistical Lives An Interdisciplinary Perspective. Oxford University Press, Oxford, 2015, p. 24-42

[Ro19] Roff, H.: The folley of trolleys: Ethical challenges and autonomous vehicles. https://www.brookings.edu/research/the-folly-of-trolleys-ethical-challenges-and-autonomous-vehicles/, accessed: June 13th, 2019

[Ta08] Taleb, N.: The Black Swan. Penguin, London, 2008

[Th17] The Moral Sense Test. http://www.moralsensetest.com/, accessed: Dec 2nd, 2017

[Th85] Thomson, J. J.: The Trolley Problem. The Yale Law Journal, Vol. 94, No. 6 (May, 1985), pp. 1395-1415

[We51] Welzel, H.: Zum Notstandsproblem. ZStW Zeitschrift für die gesamte Strafrechtswissenschaft 63 [1951, 1], p. 47-56.

Co-creating digital public services with older citizens: Challenges and opportunities

Juliane Jarke[1], Ulrike Gerhard[2] and Herbert Kubicek[3]

Abstract: Older citizens are excluded above average from digital public services as they do not meet older adults' needs and expectations. Yet most digital technologies, designed for an ageing population, reproduce particular images about age and ageing, such as the old age defined by ill health, deficits and limitations or an emphasis on active ageing. Digital public services are no different. We are interested in understanding through what kind of methods older adults may be enabled to become active co-creators of information systems and in so doing may transform our images of an ageing population. The paper is based on a collaborative research project in which older adults co-created a digital neighbourhood guide. We describe a framework of interventions which facilitated the co-creation process and discuss associated challenges and opportunities.

Keywords: older adults; co-creation; open data; participatory design; cultural probes; civic tech

1 Introduction

Interactions between public authorities and citizens are increasingly mediated by digital technologies as more and more public services are provided via digital channels. However, in many cases these services are not used widely and in particular, older citizens are excluded above average, as digital services do not meet their needs and expectations. Recently the idea of 'open government' has attracted attention, encouraging the development of so-called civic apps (digital applications that are based on open government data and developed by civil society actors such as Code4America). These civic apps are meant to provide for better and user-centred services and to foster public participation and engagement in the development and provision of public services through the use of open government data.

Older citizens—if at all—are often only marginally involved in such kind of civic technology engagement. They very rarely constitute the focal user group of civic apps; commercial web applications mainly focus on their assumed deficits and limitations (e.g. physical and cognitive decline, loneliness, dependency) [Ange16]. Hence such mediated services are predominantly based on stereotypical images of 'being old' and/or inscribe ideals of active and healthy ageing in the technology, that correspond with contemporary neoliberal concepts of optimisation and self-responsibility [Suop15, Suop16, Katz00].

[1] Institute for Information Management Bremen (ifib), University of Bremen, jjarke@ifib.de
[2] Institute for Information Management Bremen (ifib), University of Bremen, ugerhard@ifib.de
[3] Institute for Information Management Bremen (ifib), University of Bremen, kubicek@ifib.de

doi:10.18420/inf2019_08

Governments are placing an increasing emphasis on opening their data repositories so as to encourage new forms of service design and delivery [Shak13]. A growing number of cities are making their data openly available. However, such open data is normally read-only (that is, citizens are usually not able to easily suggest changes, correct errors, etc.) and there is little return for local governments [LeAW15, HuKr14]. Often developers anticipate the needs and wants of citizens based on their own experiences with lack or insufficient knowledge about prospective user groups. This is also common in other software development contexts: Often they are based on the designers' assumptions regarding older people's needs [Neve10] and are based on stereotypical images of ageing as biological and chronological process of physical, mental and social decline [WaGa18]. Hence, most information technologies designed for an ageing population are developed with the intention to counterbalance those deficiencies. In so doing they reiterate and reinforce prevalent discourses around ageing and script particular images of "age" into technology [PFJM15].

Thus, in order to create value that benefits administrations as well as citizens, it is crucial to engage citizens into the process of open data service app development, especially those who are often forgotten when it comes to technological innovations. There is a need to bring together city administrations as data owners, technology developers and older citizens as knowledgeable individuals and prospective users in order to co-create valuable public services based on open data in participatory design processes. This paper analyses and discusses challenges that emerge when co-creating digital public services for an ageing population as well as opportunities for alternative forms of civic technology. Based on an action research project in which we co-created a digital neighbourhood guide with older citizens, this paper addresses the following questions:

- How can we engage older citizens in the design of digital public services?
- What are challenges of co-creating digital public services for older citizens?

This empirical study is based on the EU-funded project MobileAge (2016-2019) in which we developed and evaluated a co-creation methodology in four European cities or regions. In this study, we report of the field work conducted in one of the co-creation sites: Bremen Osterholz.

In the following, we provide a brief theoretical introduction to the co-creation of digital public services. Subsequently we present and critically discuss our own co-creation methodology and process. We attend to the changing images of old age throughout the process. We conclude with some considerations about the roles of older adults in such a design process and reflect on the images of old age that were present throughout.

2 Review on co-creation of digital public services

The term co-creation has only recently gained attraction and is used mainly with regard to IS development and online services. In some cases, it is just another term for participatory software development or collaborative development. However, co-creation differs

in three aspects [Gomi13]: (1) Traditionally, end-users only provided information on needs and requirements and gave feedback while the experts (designers, software developers) performed the programming and design-related tasks. In co-creation, end users may also be involved in programming and design activities themselves. (2) End-users define or influence the architecture of the system, not only single features and interfaces. (3) End-users take over responsibility for the services and systems developed and may maintain (certain aspects of) it.

At the moment co-creation stands for a higher level of end-user contributions while the lower level of participatory design by occasional users has not been broadly achieved. So far, we find successful participation of citizens in the development of online services in cases where research teams play a moderating and supporting role. While participation in some co-creation initiatives is limited to co-design of the interface of an application, others also involve citizens in generating topics and contents. Participants can take different roles in the co-creation process. According to the literature [NaNa13] these roles may be: (1) Explorer: Identify problems to be solved; (2) Idea former: Generate solutions to well defined problems; (3) Designer: Design and/or develop implementable solutions; (4) Diffuser: Facilitate the adoption and diffusion of the developed solution. In our conclusion, we will also argue for a fifth role: Data editor. Generally, there is a variety of stakeholders involved in co-creation projects, covering different co-creation roles. In the following, we describe and discuss how older citizens co-created a digital district guide and how they assumed their designated roles in the process.

3 Co-creation for social participation of older adults

The focus of the co-creation activities in Bremen was on the socio-spatial aspects of social inclusion: This is a pertinent topic as the relationship and bond with their immediate living environment as well as the capability to move within a neighbourhood confidently become more and more important as people grow older. At the same time, there is an increased risk of loneliness as relatives and friends may pass away. Thus, older adults are in need to find places and opportunities for social interaction, leisure activities and civic participation, which are strongly related to their confidence of moving freely within their neighbourhood. Besides economic and health related resources and the public infrastructure, the availability of cultural and social capital significantly affects people's agency. Knowledge of the social space is an essential prerequisite for the sense of belonging, security and independence [WLGR12].

3.1 Field work in Bremen Osterholz

Bremen Osterholz is characterised by six very diverse neighbourhoods that give the district its multifaceted character. The neighbourhoods are important points of reference for the identity and movement of many people. These aspects were important to our co-creation process as the focus was on social inclusion and active participation in the district.

In May 2016 we began our co-creation process by recruiting 12 older adults living in the district. Our core-group consisted of seven women and five men; their age ranged from 55 to 80. They were comparably well educated, physically and psychologically healthy and all lived independently. Overall, the participants were familiar with digital technologies. Only one participant had never used a computer. Two participants were still employed. Almost half of participants engaged actively in political and volunteering work in the district. During our fieldwork (from May 2016 until January 2017), we conducted eight interviews with intermediaries, ten meetings with local stakeholders, two information events, 14 co-creation workshops with our core group of older adults and 12 supplementary focus groups with about 80 older adults in the district. In addition, 11 of our core participants completed a set of probes (for a description of the method see below) and were interviewed individually. We used an action research approach that included a reflective learning methodology. Most interviews and focus groups were audio recorded and transcribed. Each intervention was documented with respect to its date, length, participants, objectives, activities, observation notes, and learning outcomes. As we understand co-creation as a reflective practice, we reflected upon our interventions and adjusted the process continuously. These adjustments were also documented. The service idea that was developed, refined and implemented throughout the co-creation process was a digital, interactive neighbourhood guide.

In the following, we describe the co-creation process. We conclude with implications from our experiences for future co-creation processes with older adults. In doing so, we also reflect on the ways in which particular images about old age evolved in the process and how these images facilitated or contested the stereotypical images of older adults inscribed in 'traditional' software development. What was striking was that the participants' self-perception as being old is essentially constructed through their differentiation from others; in particular, through the ways our participants understood themselves as trailblazers for future older citizens.

3.2 Processes of co-creation

Planning and evaluating co-creation

The planning and evaluation of co-creation interventions is a continuous process throughout. Key to co-creation as a successful reflective practice is the continuous documentation of co-creation interventions through a *protocol of activities* and/or a *reflection journal*. Questions to be considered for these types of documentation concern the aims of an intervention, their implementation and their outcomes; reflections on their effectiveness and appropriateness as well as corrective actions.

Another important stream of activity is the planning of the evaluation of co-creation interventions. From the start of the project, the core team should define *what* will be evaluated (e.g. process, product) and *how*. The evaluation is, of course, very much dependent on the context and domain in which co-creation activities take place and may be refined as the co-creation activities proceed.

Recruiting and engaging stakeholders

The engagement with stakeholders and the recruitment of co-creators proved to be a continuous activity throughout the process. While ideas developed, the service concept became more refined and the required data were defined and collected, complementary focus groups and engagement with additional local stakeholders (such as service providers or data owners) were required. However, the recruitment of the core group was mainly conducted via newspaper articles, were we addressed older adults in the district that were knowledgeable and/or interested in their district. Although we explicitly addressed people with and without experience with digital technologies, most of the participants were already using smartphones, PCs, tablets and/or the internet. All of them shared an interest in these 'new# media technologies.

To start the co-creation process we wanted to provide a notion of the project's objective and what kind of input, in particular local knowledge we would like participants to contribute. As these expectations are difficult to communicate verbally, we decided to begin the process with something tangible: An activity that would be fun and attract interest in the project, so that people would be encouraged to come again. We choose to develop a card game in order to (i) learn about the district, (ii) facilitate the communication between participants and (iii) provide low-tech engagement. At an *information event* participants were asked to fill out questions on the cards which related to their district. In doing so, they not only shared their knowledge about the district (e.g. what is beautiful in Bremen Osterholz) but also considered questions that could be relevant to others in the district. For a *kick-off workshop* we prepared a proper card game (with pictures) based on the participants' input. Their task at this workshop was to evaluate each other's input via blue and green points (for relevance) and leave remarks. For our process, it was important to establish the co-creators as experts of the district and to appreciate their local knowledge. This established an engagement of mutual respect between the project team and participants, as both parties wanted to learn from the other.

The participants appreciated the refined version of the card game, as they could see that their work had been valuable and were actively engaged with the card game. To see pictures of their district and discuss them seemed to motivate them. The card game as a method worked well to motivate the participants as the focus was laid on the district, not on technology. It enabled them to form a sense of community based on their shared practices of living and ageing in the district. At the same time, in interacting with the card game a shared notion of what it means to grow old in the district emerged that comprised of being interested in and care for the local environment and its people, being knowledgeable of people, places and institutions and being actively involved in the social and cultural life in the district.

Co-creating a service concept

The initial tasks associated with the co-creation of a service concept included a preliminary survey and analysis of existing services as well as the development of first ideas. The service to be developed was defined in the co-creation process, but we had to have a

concrete idea about

- What service domain we developed a service for;
- What the thematic space of the service was;
- Who the target user group was and what other stakeholders were relevant;
- What kind of technical solution was going to be developed (mobile app, website).

In order to address these kinds of questions, we had to understand the everyday practices of older people in the district better. To understand what it means to age in this particular place. While the card game offered a first interaction with our participants, there was a need to explore and learn about their everyday lives in a more structured way. For this reason, we developed a set of 'cultural probes' [GaDP99, BoGB12, JaMa18] which are descriptive and exploratory tasks that are (typically) based on self-reporting. In our case, the participants kept the probes for 10 days. They collected data on themselves, their lives and their socio-spatial and media use practices. Follow-up interviews were conducted individually to prepare and accompany the process and a de-briefing session (workshop) to supplement, validate and explore the data.

In contrast to more traditional approaches to probes which are used in user-centred design [SaSt08], probes in our project were used as a method and tool for co-creation. Hence, in addition to their inspirational function and tool for the requirement elicitation, we also used the probes as a communication and engagement tool for the subsequent co-creation process. In a follow-up workshop, the participants jointly reflected on the activity and their experience. The aim was to (1) jointly reflect on the probes activity and experience and to (2) identify some key characteristics that defined their everyday practices in the district.

One task concerned a neighbourhood map. The main aim of this probe was to understand social inclusion with respect to primary networks and space. Participants were asked to highlight where they live (red dot), where friends & family live (blue dots), where important places for their everyday are (yellow dots). What we were interested in learning from this map concerned, for example, how connected our participants felt to people/places and the spatial dimension of their primary networks (neighbourhood, quarter, district, clubs). We were also interested in learning which social networks the participants were part of and where they meet. The returned maps differed greatly with respect to the extent of the networks and the mobility patterns. The maps were supplemented with diaries and a set of seven maps in which participants documented their routes for a week.

When participants compared the individual maps during the workshop, they discussed what they believed to be differences and characteristics. Some of the key differences where: biographical (on whether somebody just recently moved to Bremen Osterholz), related to retirement/employment, living circumstances (alone vs. partnership vs. caring for partner) related to mobility & functional health, related to the financial situation and

how active people were in terms of charity work and hobbies. All these considerations were noted and informed the subsequent development of 'personas' and 'scenarios' [RoCa02, Carr00]. Importantly these 'personas' were not 'prototypical users' grounded in the stereotypes of software developers but were rather characters that were defined by characteristics deemed important and relevant to our participants with respect to ageing in in this district. For example, the biological age or gender did not play a role for identifying differences across the socio-spatial networks depicted in the map. Our participants rather pointed to specific place-making practices that resulted in the different maps.

For the participants the probes facilitated an awareness of everyday practices and practices related to the appropriation of the district when becoming older. They sensitised participants about certain aspects of their everyday practices and were hence tremendously helpful in identifying needs as well as resources. For the researchers they allowed to develop a better and more profound understanding of these practices. This demonstrated that probes were superior to interviews in which participants could, for most parts, only report on their everyday live without prior reflection.

The probes solidified the image of the older participants as being tightly connected in their neighbourhoods and strongly engaged with their socio-spatial environment. In addition, our participants put an emphasis on "being active", e.g. the everyday diaries were full of activities; times of loafing almost did not appear. Furthermore, some participants were motivated to improve their physical exercise through the documentation of their mobility patterns [for a more detailed discussion on the use of probes in this project see JaGe17].

In the subsequent workshops, the personas provided a good basis to discover and discuss the information needs of the older citizens. They were helpful in order to encourage participants to think beyond their own wishes and needs and to relate to others who might be different from them. Furthermore, they allowed the participants to address sensitive issues by referring to a third person. Importantly, the personas were not developed through stereotypical ideas about older adults but rather in collaboration with them.

Overall, the result was a manifold of relevant object categories and attributes to be visualised on the map, which later turned out to be too numerous for the scope of the project. Further, the personas helped to generate ideas for the service definition. The main point here was that the participants felt that it was important to focus on the resources an older person has: They told us how they were helping friends, relatives and neighbours for example support in housekeeping or getting somewhere. Here it became eminent how the participants experienced and represented themselves as efficacious with respect to themselves and to others. One idea for a service was to support the exchange of time, goods, or abilities. These considerations were in stark contrast to most of the service developed for older adults that centre around their deficits and aim to support for example, health-related support service.

As part of the service and data definition, we held two further workshops: one on the informational content and one on interactive elements of the MobileAge app. The aim of

these workshops was to select the categories of objects to be shown on the map, to determine attributes for each category of objects and further to define the relevant information about these objects. During the workshop, we divided the participants in groups of 2 – 3 to work on different categories of objects. We had prepared lists of objects per category. As we were interested in considering what kind of information would be interesting about the objects, we had also provided supplementary information in form of leaflets and Websites print-outs to the groups. The workshop concluded with presentations and discussions of the results.

In a subsequent workshop we decided with the participants to develop a map-based service. We agreed that only a limited number of categories of objects could be included in the neighbourhood guide as only very limited data was available and hence an intensive data creation process was ahead of us. The decision was supported by the argument to focus on those categories of objects that are not currently systematically captured anywhere else (e.g. nice places, informal meeting places). This would constitute an added value, particularly with regard to the content (as making available informal local knowledge).

Working with (open) data

One of the first steps was to generate a report about the data that were available for our topic and determine how appropriate these were. Subsequently the stream of activity led to the collection and validation of data that were identified as relevant but were not yet open or needed to be collected across various data owners. A further activity included the creation and integration of new (open) data by the core project group and co-creating older citizens. Lastly, the service and collected data had to be presented in a meaningful way to users. Editorial work (such as descriptions about data objects) was necessary.

In order to start the co-creation of data on the selected categories of objects and respective attributes we created a matrix (table), selected the respective institutions included in a district reader and filled the table. Only data on a few attributes could be matched with available open government data, e.g. public benches close to nice places. Data for most of the attributes had to be specified and collected. For this purpose, we arranged different focus groups to amend and complete the list of institutions and data on their attributes. The data tables with attributes were central to our co-creation activities, with most activities providing input for their structure, completion, validation and subsequently visualisation.

According to the selection of categories of objects and relevant attributes, we decided to differentiate between two main kinds of objects, with differing attributes:

- Nice places and walks, with descriptions about what was considered particularly nice, and information about the availability of benches and toilets nearby as well as supplementary information on possibilities for exercising and BBQ.

- Informal meeting facilities, institutions and services in the field of culture, consultancy and advice as well as sports with data on the individual services and facilities, events, contact person etc.

For each object, we created a matrix with a line for each object and several columns for the different attributes. These two data tables became the central working tool for the data collection and co-creation process with two objectives: (1) *Completeness*, e.g. identify all the relevant objects in Bremen Osterholz for each category; (2) *Richness of relevant details*, e.g. to collect data on as many aspects as possible for each object. All the interventions mentioned above served these two purposes and gradually completed the tables. While information on attributes such as address, contact, website was evident and easy to collect, the description was the most difficult one. The purpose of the description was to communicate why a place is nice or a facility of interest to older people. For the description, our core group participants mainly had contributed keywords. In order to acquire this information, participants assumed responsibility for particular objects (e.g. places), validated the information (e.g. through going there) and creating data (e.g. photographs).

Overall, we had to realise that very little open government data was available on the content identified as most relevant by our participants (social, cultural, leisure activities). Some participants engaged heavily in collecting data, while others were happy to name objects of interest but not to collect or validate detailed data on attributes.

Co-creating software

The visual design and functionality of the app were co-created through a number of paper-prototyping exercises and slowly transformed into digital prototyping. A first step for the co-creation of software was to identify concepts and app ideas, then gather requirements from each stakeholder. These ideas became more refined as the service co-creation activities proceeded and relevant data sets were identified (and created). The stream of activities concluded with the testing and reviewing of the functionality.

In order to enable members of our core group to test the application prototype and to validate and complete the information, we provided the participants with tablets. In a workshop, we gave nine tablets to the members of our core group. The participants kept the tablets for eight weeks. They received a short introduction on how to use the devices and how to test the first prototype.

In the observations of their use practices and a focus group around the tablet use, we developed a more pronounced understanding of the participants' motivations to appropriate certain "new" media technologies. Our participants' overall curious attitude towards new media technologies was not primarily rooted in an enthusiasm for these technologies themselves. Rather they shared a self-perception of socially engaged and politically interested citizens and they were aware of the growing importance of the internet and digital devices for society at large and social relations, in particular. In order to be able to fully participate in today's society they felt the need (and to some extent social

pressure) to keep up with these technological developments. In this regard our participants perceived themselves as pioneers/trailblazers in their generation and felt a sense of responsibility to convince "off liners" to start using mobile devices and the internet (i.e. by showing funny YouTube videos on the smartphone).

In particular, those participants who only had a desktop computer and no mobile device appreciated the opportunity to test a tablet not only for the purpose of our project but beyond. The introduction of tablets and the opportunity to test the co-created website was an important step in the process. Besides the experience to use a tablet, they could experience how their efforts and input had been integrated and valued.

Regarding the technical solution, it was necessary to consider the technological infrastructure available in the specific area. This included internet coverage as well as the supply of devices. Furthermore, the engagement with technology among the concerned older population had to be taken into account. This was partly done by reviewing statistics/studies on technological infrastructure and access for the particular region/area.

The city district guide for older citizens had to meet several requirements with regard to content and technical functions. With respect to content, the relevant objects had to be covered as comprehensibly as possible, e.g. all existing places and meeting points with all the relevant attributes. With regard to functionality, it had to be easy to find these objects. To meet these two requirements, different competences in the project team were required as there are in professional app development (e.g. for content, functionality, design). While for some design questions it was appropriate to present different existing websites, for other aspects paper prototyping was more adequate. It turned out that the exercise with an open screen and several paper elements for possible menus, left room for discussion of many associated issues. This exercise only came to a result once the researchers intervened and moderated the discussion. While some participants enjoyed the paper prototyping others were hesitant to „glue" their proposition on paper. For those who were not too acquainted with digital media the design task appeared to be too tedious. For those that regularly used digital media the ideas about design were mainly derived from their own experience with existing websites and applications.

Exploiting the service

For the initial planning of co-creation activities, a first definition of targets, outputs and value propositions was defined. This also included initial considerations about the sustainable deployment of the service and its required data and technical infrastructure. Subsequently we developed ideas on how the service might be maintained beyond the end of the project. We agreed that the city information portal would maintain the app and technical aspects, whereas a group of local stakeholders would be responsible in maintaining the content.

4 Conclusion

4.1 Implications for engagement

Overall the recruitment and engagement mainly involved already active senior citizens (e.g. through computer clubs or charity work). Participants may be actively involved as explorers and idea formers, but the degree of participation decreases for design activities, and may increase again for diffusion. The older co-creators mainly selected from a number of given alternatives, and to some extent selected from self-defined options. Overall, participants' needed to find their new role from a customer/user of a service to a service designer/co-creator. Regarding older citizens' possible roles in a co-creation process, our experiences demonstrate that the role model proposed in traditional participatory design projects focuses too much on technology design and disregards the co-creation of the content of a service.

Relevant data are not provided by a single open source but are distributed across various stakeholders and organisations or are not open nor available at all. Often there is no quality assurance of data sets. Hence, data sources have to be investigated and validated thoroughly. In the few cases where open data are used, the data are provided by authorities already involved in the projects (e.g. public transportation) and the participants are not involved in the creation of data as was suggested by others. Rather, one of the key tasks of participatory open data projects seems to be the co-creation and opening of data sets. A new role for co-creators in such projects could then be the role of data collector and creator.

In projects driven by the administration or in data-driven projects, where the content is clear and the aim of the co-creation process focuses on the development of a technical system to distribute certain available data, these roles might be adequate. However, in a citizen driven co-creation process where the citizens define the services to be developed, the task of data co-creation needs to be added. We therefore extend this model with the role of a data editor, which comprises the tasks of data definition, collection, creation, integration, validation and maintenance.

4.2 Implications for innovating co-creation

The involvement of older adults in processes of developing new services is most essential and fruitful in the beginning of the co-creation processes when the service to be developed is defined. In this early phase, the sphere of influence is biggest and it is decided whether the outcome will be meaningful and have an impact. Furthermore, in this phase no technical skills and competencies are necessary whereby potential barriers for participation are reduced. However, the early involvement of older adults in the co-creation process has some implications with respect 1) to the recruitment of participants and 2) to the procurement of data.

1) In order to provide for substantial participation, the result of a co-creation project

needs to some extend be left open in the beginning. On the other hand, it is important to explicate expectations, duties and tasks when recruiting.

2) When involving older adults in the definition of the service to be developed, service providers cannot rely solely on open government data. Since the needs, wishes and interests of future users might not be met by the data available, service providers and developers need to consider other data sources as well as the co-creation of data. So far, the task of data procurement has been underestimated: In open data projects, there was no need for it because of a supply-driven approach to service innovation.

4.3 Implication for our understanding of old age

While co-creating the digital neighbourhood guide "for and with older adults" we had to negotiate with our participants - explicitly and implicitly - what it means to grow old. In the different phases of the process, the participants positioned themselves with regard to their neighbourhood, technology and other people, and thereby created their own images of age and ageing. The main characteristics of these images comprise of the outstanding importance of the local environment, different forms of being active (physically as well as politically or culturally) and keeping up with current (technological) developments. With regard to all of these characteristics, our participants viewed themselves as different to other older adults and thereby formed a shared identity as being old/ageing differently to others: They felt that they are more interested and knowledgeable than others about the district, more actively engaged and more open-minded towards new media technologies. Not surprisingly, we can relate these images to current dominant discourses of 'successful ageing' and the accompanying alternative images of ageing as a process of physical, mental and social decline. It is noticeable that all participants strive to fill their daily lives with as many and varied activities as possible. This applies to physical and mental activities as well as active engagement in social, political or cultural areas and came to be represented in the types of objects displayed in our digital district guide (nice places, meeting places, culture). The active and self-responsible work on one's own body and mind, as well as the effort to keep-up with current developments and to fill one's spare time - after many years of professional and/or domestic duty - with honorary and non-profit activities, also point to current discourses on 'active' or 'successful' ageing. The ideals of lifelong productivity and self-management are often associated and critically questioned in gerontological research with current trends such as the dismantling of the welfare state and the demand for greater self-responsibility. In this perspective, we find that the co-creation approach does not automatically lead to the contestation of hegemonic discourses and related stereotypes. Since the images of ageing and old age co-constructed in the co-creation process described, heavily influenced the service defined and the software developed, it remains to be seen, how older adults actually adopt applications such as the digital neighbourhood guide, how and with which consequences such technologies are used.

5 Funding

MobileAge has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 693319.

6 Acknowledgments

This co-creation project would not have been possible without the support of the many older residents, the local district council and local social care service providers in Bremen Osterholz. We would like to thank the co-creators of our project.

Bibliography

[Ange16] ANGELETOU, ANGELIKI: *Designing tech for Old People; Who's Old?* URL https://medium.com/@angel_alice/designing-tech-for-old-people-whos-old-1743d7cd538b#.dajjtf7i9

[BoGB12] BOEHNER, KIRSTEN ; GAVER, BILL ; BOUCHER, ANDY: Probes. In: LURY, C. ; WAKEFORD, N. (Hrsg.): *Inventive methods: the happening of the social*. London : Routledge, 2012 — ISBN 978-0-415-72110-3, S. 185–201

[Carr00] CARROLL, JOHN M.: *Making Use: Scenario-Based Design of Human-Computer Interactions* : MIT Press, 2000

[GaDP99] GAVER, BILL ; DUNNE, TONY ; PACENTI, ELENA: Design: Cultural Probes. In: *interactions* Bd. 6 (1999), Nr. 1, S. 21–29

[Gomi13] GOMILLION, DAVID: *The Co-Creation of Information Systems* : Florida State University, 2013

[HuKr14] HUNNIUS, SIRKO ; KRIEGER, BERNHARD: The Social Shaping of Open Data Through Administrative Processes. In: *Proceedings of The International Symposium on Open Collaboration*, *OpenSym '14*. New York, NY, USA : ACM, 2014 — ISBN 978-1-4503-3016-9, S. 16:1–16:5

[JaGe17] JARKE, JULIANE ; GERHARD, ULRIKE: Using cultural probes for co-creating a digital neighbourhood guide with and for older adults. In: *Mensch und Computer 2017-Workshopband* Bd. 93 (2017), S. 79–85

[JaMa18] JARKE, JULIANE ; MAAß, SUSANNE: Probes as Participatory Design Practice. In: *i-com* Bd. 17 (2018), Nr. 2, S. 99–102

[Katz00] KATZ, STEPHEN: Busy Bodies: Activity, aging, and the management of everyday life. In: *Journal of Aging Studies* Bd. 14 (2000), Nr. 2, S. 135–152

[LeAW15] LEE, MELISSA ; ALMIRALL, ESTEVE ; WAREHAM, JONATHAN: Open Data and Civic Apps: First-generation Failures, Second-generation Improvements. In: *Commun. ACM* Bd. 59 (2015), Nr. 1, S. 82–89

[NaNa13] NAMBISAN, SATISH ; NAMBISAN, PRIYA: *Engaging Citizens in Co-Creation in Public Services Lessons Learned and Best Practices*. Washington, DC : IBM Center for The Business of Government, 2013

[NBRS19] NEATE, TIMOTHY ; BOURAZERI, AIKATERINI ; ROPER, ABI ; STUMPF, SIMONE ; WILSON, STEPHANIE: Co-Created Personas: Engaging and Empowering Users with Diverse Needs Within the Design Process. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems - CHI '19*. Glasgow, Scotland Uk : ACM Press, 2019 — ISBN 978-1-4503-5970-2, S. 1–12

[Neve10] NEVEN, LOUIS: 'But obviously not for me': robots, laboratories and the defiant identity of elder test users. In: *Sociology of Health & Illness* Bd. 32 (2010), Nr. 2, S. 335–347

[ÖOJF15] ÖSTLUND, BRITT ; OLANDER, ELIN ; JONSSON, OSKAR ; FRENNERT, SUSANNE: STS-inspired design to meet the challenges of modern aging. Welfare technology as a tool to promote user driven innovations or another way to keep older users hostage? In: *Technological Forecasting and Social Change* Bd. 93 (2015), S. 82–90

[PFJM15] PEINE, ALEXANDER ; FAULKNER, ALEX ; JÆGER, BIRGIT ; MOORS, ELLEN: Science, technology and the 'grand challenge' of ageing—Understanding the socio-material constitution of later life. In: *Technological Forecasting and Social Change* Bd. 93 (2015), S. 1–9

[RoCa02] ROSSON, MARY BETH ; CARROLL, JOHN M.: *Usability Engineering : Scenario-based Development of Human-computer Interaction*, *Morgan Kaufmann Series in Interactive Technologies*. Bd. 1st ed. San Fancisco : Morgan Kaufmann, 2002 — ISBN 978-1-55860-712-5

[SaSt08] SANDERS, ELIZABETH B.-N. ; STAPPERS, PIETER JAN: Co-creation and the new landscapes of design. In: *CoDesign* Bd. 4 (2008), Nr. 1, S. 5–18

[Shak13] SHAKESPEARE, STEPHAN: *Shakespeare Review: An independent review of public sector information*. London : Department for Businesses, Innovation & Skills, 2013

[SiJo15] SIEBER, RENEE E. ; JOHNSON, PETER A.: Civic open data at a crossroads: Dominant models and current challenges. In: *Government Information Quarterly* (2015)

[Suop15] SUOPAJÄRVI, TIINA: Past experiences, current practices and future design. In: *Technological Forecasting and Social Change* Bd. 93 (2015), S. 112–123

[Suop16] SUOPAJÄRVI, TIINA: Knowledge-making on 'ageing in a smart city as socio-material power dynamics of participatory action research. In: *Action Research* (2016)

[VPWO15] VINES, JOHN ; PRITCHARD, GARY ; WRIGHT, PETER ; OLIVIER, PATRICK ; BRITTAIN, KATIE: An Age-Old Problem: Examining the Discourses of Ageing in HCI and Strategies for Future Research. In: *ACM Transactions on Computer-Human Interaction* Bd. 22 (2015), Nr. 1, S. 1–27

[WaGa18] WANKA, ANNA ; GALLISTL, VERA: Doing Age in a Digitized World—A Material Praxeology of Aging With Technology. In: *Frontiers in Sociology* Bd. 3 (2018)

[WLGR12] WILES, JANINE L. ; LEIBING, ANNETTE ; GUBERMAN, NANCY ; REEVE, JEANNE ; ALLEN, RUTH E. S.: The Meaning of „Aging in Place" to Older People. In: *The Gerontologist* Bd. 52 (2012), Nr. 3, S. 357–366

IT-supported Hospital Discharge Management – Findings of a Multi-Method Research Design

Sarah-Sabrina Kortekamp [1] Maria Carmen Isabel Süßmuth [2] Andrea Hildner [3] Ingmar Ickerott [4] Frank Teuteberg [5]

Abstract: A structured hospital discharge management process can lead to a smoother transition to aftercare. In practice, providing continuity of nursing care after a stationary hospital stay is accompanied by numerous challenges. The presented study aims to point out the use cases and requirements for an IT system supporting the diverse tasks of the participating actors. Within the scope of a multi-method research design, the authors conducted and analysed stakeholder interviews, a shadowing, a systematic literature search and statutes in order to gain the presented results. This publication presents 37 requirements, grouped to 14 use cases. A process model in BPMN visualises the discharge management process. Further, the authors derived implications for practice and research. These can be used for the development, classification and assessment of IT systems. Therefore, this publication provides a significant contribution to the development of socio-technical systems within the health care domain.

Keywords: Discharge management, use cases, requirements, hospital information system, multi-method research design

1 Introduction

The German health care system is confronted with multiple problems. One of the central subjects of social policy is to ensure the sustainable protection of health. Steadily rising costs, the uncertain financing situation, the lack of qualified personnel, the decrease of the nursing care potential within family structures as well as a fluctuating care quality constitute enormous challenges [PG12].

The intersectoral sharing of patient data is particularly complicated regarding the transition between inpatient and outpatient health care facilities and linked to media disruptions. As a consequence, frequent mistakes and information losses originate that can lead to serious health complications for the patients [MS14; WH16]. When applied to the discharge

[1] Hochschule Osnabrück, sarah.kortekamp@hs-osnabrueck.de
[2] Hochschule Osnabrück, m.suessmuth@hs-osnabrueck.de
[3] Gesundheitsregion EUREGIO, andrea.hildner@gesundheitsregion-euregio.eu
[4] Hochschule Osnabrück, i.ickerott@hs-osnabrueck.de
[5] Universität Osnabrück, frank.teuteberg@uni-osnabrueck.de

doi:10.18420/inf2019_09

management as the central topic, the following specific problem situations can be derived [Ja17]:

- A lack of capacities in aftercare facilities (e. g. short-term care, rehabilitation, ambulatory care).
- The excessive administrative burdens imposed by health insurances.
- Undefined processes within the existing health care standards, in particular regarding the interfaces to other occupational groups and institutions.
- The missing of standardised IT structures for the intersectoral communication between the health care actors.

The expert standard discharge management in healthcare [Ex09] shines a light on some of these problems and also their solutions. Although it is highly acknowledged, most facilities implement the standard only in part, leaving out important details like the reevaluation of the process after the discharge of a patient.

Accordingly, the legislator has created a reorganisation of the law (GKV-Versorgungsstärkungsgesetz) in §39 Abs.1a SGB V to strengthen the provision in the legal health care binding regulations for discharge management. This led to the framework contract discharge management [Ra16] as an agreement between the German Hospital Association, the German National Association of Statutory Health Insurance Physicians and the "GKV-Spitzenverband".

From 1st of October 2017, hospitals are obligated to organise the discharge management for patients after a full- or part-stationary stay or who have received similar services. The discharge management is the central component of high-quality treatment and shall guarantee a continuous provision of care for patients after their discharge from the hospital.

Discharge management aims to plan and ensure an unproblematic transfer for patients where post-hospital care can otherwise not be guaranteed. Here, the influencing factors are the patients' state of health, the social environment as well as the domestic situation and the financial opportunities. The care can be provided in various ways and is oriented towards the individual needs and preferences of the patients. Common follow-up care facilities are rehab hospitals, residential care homes for the elderly and ambulatory care.

Introducing discharge management requires not only a reorganisation of processes but should also encourage a change of roles and tasks and thus lead to a more interdisciplinary workstyle. Only if this preliminary work is completed, an implemented IT system can support the discharge management in a useful manner [DM12]. In this context, the following research question arises:

What use cases can the IT-system support for a hospital discharge management and what requirements for such a system exist?

This paper is divided into five sections. Section 2 describes the chosen study design and applied methods. Section 3 contains two subsections. The first subsection presents the identified use cases and corresponding requirements, whereas the second visualises and describes the use case dependencies based on the presented process model. In the ensuing section 4, the authors discuss the presented results. In this context, practical and theoretical implications are given and limitations of the study are pointed out. The last section presents the conclusion.

2 Method

2.1 Multi-method research design

The study design follows the procedure of requirements engineering described by Sommerville [So05]. He divides the process into the phases Elicitation, Analysis, Validation, Negotiation, Documentation and Management. The requirements often change during the development of a system, so that the collection and consolidation should reflect an iterative process. This change is due to several factors. Among others, initially stated requirements are often rather vague, which might lead to different interpretations from the developers' and users' perspectives. Therefore requirements should be surveyed early on. The developing team should discuss and refine them regularly. Early low-fidelity mockups can help throughout the process. The following shows a short overview of our approach:

Elicitation: Data collection including interviews and literature search.
Analysis: Analysis of the source material by means of inductive coding [Ma14] and review of the identified use cases and requirements regarding conflicts and overlaps. Discrepancies were discussed and a list of requirements was deducted.
Validation: Presentation of the requirements to the stakeholders with ensuing disucssion.
Negotiation: Discussion of unclarities and conflicts with the stakeholders until consens was achieved.
Documentation: Development of a BPMN process model.
Management: The requirements will be updated regularly.

By combining practical experiences of health care professionals with evidence from literature, we formed a holistic view of the requirements. This approach is qualified for understanding the high complexity and multi-dimensionality of real-world problem situations [De73].

2.2 Semistructured interviews

During March 2016 and January 2017 we conducted six interviews, three of them in a hospital (discharge-management worker, head of the central-patient admission, head of the gerontology ward), two with the management of follow-up care facilities and one with two workers at the care support point.

The interviews aimed to survey the current healthcare situation and unsystematic processes within a rural region. In order to establish comparability between the interviews, we used an unstructured interview guideline. The topics included (1) the current and future state of the healthcare, (2) processes and problems of the discharge management as well as (3) the intersectoral communication within the healthcare system.

In order to generate unbiased results, the interviewer recorded the interviews with a dictation machine, and professionals conducted a literal transcription. Afterwards, the authors used the transcription to analyse the interview contents using the method of 'inductive coding' [Ma14].

2.3 Shadowing

In addition to the interviews, a shadowing of a hospital discharge management worker was carried out by two researchers. The aim was to generate a more thorough understanding of the necessary tasks and processes regarding discharge management. The researchers behaved in such a way as to minimise the disturbance of the shadowed worker. The hospital staff and other participating people were asked to ignore the researchers while fulfilling their everyday tasks. In order to ensure an unbiased analysis, the observing researchers were not involved in the interviews beforehand.

The shadowing was documented through a defined protocol for note taking. Following the shadowing, the observing researchers first compared their notes and discussed deviations. In a second step, they consolidated the protocols. Subsequently, the researchers applied the same method of 'inductive coding' [Ma14] to derive the requirements.

2.4 Systematic literature search

The authors conducted a systematic literature search following the instructions by Webster and Watson [WW02]. For this purpose, we searched EBSCOhost and GoogleScholar (title only) for the term "discharge management OR Entlassungsmanagement OR Entlassmanagement". In order to ensure a high-quality standard, the search was narrowed down to peer-reviewed research papers, published between January 2011 and June 2018.

The search generated 291 possible relevant publications (EBSCOhost: 101 results; GoogleScholar: 190 results). First, two researchers analysed all titles and abstracts for relevance, resulting in 40 relevant publications. Second, the two researchers revised the full texts of the remaining 40 publications which lead to 14 publications that are relevant to the topic in discussion.

The authors then analysed the content of the 14 relevant publications regarding requirements for hospital discharge management. Similar to the analysis of the interviews and shadowing, we used inductive coding [Ma14] to extract the requirements.

In addition to the systematic literature search, statutory requirements regarding the discharge management as well as the expert standard discharge management in healthcare [Ex09] were analysed.

3 Analysis and Findings

3.1 Consolidated requirements and use cases

Use cases are qualified for a clear presentation of functional requirements. They provide a tool for describing the tasks and goals of the involved actors as well as the desired system behaviour in different situations. The result can be used in order to debate the level of the system development in a group and to present the planning to an interdisciplinary team of stakeholders [Co08]. We used the principles described in [Co08] to gain short and meaningful use cases.

We identified a total of 37 requirements based on the semistructured interviews, the shadowing and the systematic literature search. In order to provide a clear and more helpful overview, we grouped the requirements with similar traits and used this as a basis for the development of the use cases. Eleven requirements are valid for all interactions with the software. They were assigned to the three inductively formed categories: *documentation and input of patient data*, *interoperability* and *security* (cf. Table 1).

Category	Requirements
Interoperability	Adapts the terminology depending on the user's profession [c,[Hü15; MS14]]
	Provides a user and rights management [c,[Ra16]]
	Is accessible by different users at the same time [a,b,c,[Ra16]]
	Enables all authorised staff to access relevant patient data [a,c,[Hä17]]
	Integrates seamlessly in all work processes [b,c,[Hü15]]
	Supports standardised processes and responsibilities [c,[Hä17; HK13; MS14; PK11]]
Documentation and input of patient data	Runs a real-time validity check during data input [b,c,[Ra16]]
	Provides help functionality for filling out predefined forms and reports [b,c,[Ra16]]
	Is able to detect errors during data input [c,[MS14; Ra16; WH16]]
Security	Supports the verification of sent and received electronic documents [b,c,[Ra16]]
	Supports high data security standards [a,c,[Ra16]]

[a] *Interviews;* [b] *Shadowing;* [c] *Literature*

Tab. 1: Superordinate requirements

The requirements in the category named *Interoperability* constitute the system's basis. Due to the broad range of different electronic hospital information systems, it is essential that the described system is compatible with the existing one when it comes to terms of data input and retrieval. In this way, all relevant actors can access and edit the patient's data.

Directly linked to the first category, is the category *Documentation and input of patient data*. The system supports the input and modification of data. Portability ensures that data can be input where it is collected. The prompt input of data diminishes the chance of incomplete and outdated information. To further decrease the chance of faulty data, the IT system should be able to check the entered information for validity and inconsistencies. Also, data can be used to fill in, send and export formulas automatically. Manually exchanging data and redundant data sets were a common issue for the interviewees. For example, the process of patient admission to patient discharge at the hospital involved five data transitions from digital to non-digital data and vice versa regarding a patient's medication information.

The last category *Security* holds a particular position in the management of health care data. It is not only necessary to deny unauthorised persons access to the data but also to verify the consignor.

Table 2 shows the remaining 26 requirements on the left and the derived use cases on the right. Each requirement regarding the discharge management of a patient that involves interaction with the software is represented in one of fourteen use cases. They are numbered consecutively for cross-referencing in the text and figure 1.

One critical factor in these use cases is that users can directly enter new or missing patient data in the system (1). Due to low capacities at follow-up care facilities, the discharge management worker has to plan a patient's hospital discharge as soon as possible. Therefore, the identification of vulnerable patients is essential (2). On the one hand, this ensures early detection of patients requiring discharge management and on the other hand, no resources are wasted on patients not requiring discharge management. In doing so, an IT-supported systematic assessment should be implemented, and the reporting of eligible patients must be as easy as one click (3).

In order to not miss important deadlines (e. g. for cost refunds) or guarantee a seamless transition to the new care or rehab facility, it is the time component that matters. In no case should the IT-system hinder the discharge management worker but instead support and automatise tasks. Therefore, the IT system helps by assigning reported patients to the responding worker and providing helpful overviews regarding patients, tasks and deadlines (4). Additionally, it provides an overview of possible follow-up care facilities and medical aids (8, 9). Further support is given, by providing an easy and fast way to document the whole process as well as exporting formulas for applications and accounting (7, 10).

The discharge management worker has direct contact to the patient in order to access the pre-hospital living situation as well as wishes regarding post-hospital health care, e. g. statutory or ambulant care and the favoured care provider (5, 6).

Also, the actors are subject to considerable constraints regarding their time management. Receiving and sending messages without the need to wait for their counterpart saves valuable time. Time is also a factor when it comes to the electronic exchange of patient information and data (11-14). Being able to send data right from the IT system to follow-up care facilities

	Use Cases	Requirements
1	Input patient data into IT system	Supports input and storage of all necessary patient data [a,b,c,[Ra16]]
2	Assess critical patients	Supports a systematic vulnerability assessment of new patients [a]
3	Report critical patients	Supports a one-click report of critical patients [a] Inhibits redundant reports of critical patients [a,b]
4	Plan tasks	Assigns reported patients automatically to the specified responsible worker, e. g. based on the patient's ward [a] Gives an overview of reported critical patients incl. sorting functions [a,b,c,[Br13; Ra16]] Supports full calendar functionality [a] Shows tasks in a calendar view [a] Shows an overview of all due tasks [b,c,[Ex09; Ra16]] Supports reminders and priority settings for scheduled events [a,b,c,[Br13; Hä15; HK13; Ra16]] Shows the responsible discharge management worker [b,c,[Ra16]]
5	Obtain informed consent from patient	Allows the documentation of the informed consent [c,[Hä15; Hä17; KS16; MS14]]
6	Assess patients needs and wishes	Allows simple input of data acquired during patient interviews on-the-go [a,c,[Ex09; Ra16]] Must be portable [b] Allows to search and access structured data and application required for the cost unit [a,b,c,[Ra16]]
7	Request refund at cost unit	Supports the documentation, accounting data and appication required for the cost unit [a,b,c,[Ra16]]
8	Request medical aids	Shows information regarding further treatment possibilities [a]
9	Look for free capacities of follow-up care facilities	Provides an interactive overview of possible follow-up care facilities and home care [a,b,c,[MS14; Ra16]] Provides an overview function showing free capacities from follow-up care providers [a,b]
10	Draw up and send statement of account to cost unit	Generates documents such as transfer sheets from saved patient data [a,b]
11-14	Communication with follow-up care provider/facility	Is linked to all necessary stakeholders such as transportation [a] Enables the exchange of data and information with all relevant follow-up care givers [a,b,c,[Br15; Ex09; Hä12; HK13; Me17; MS14; Os15; Ro13]] Is able to exchange electronic documents with all relevant follow-up care givers [a] Enables internal and external synchronous communication [b] Enables asynchronous communication with stationary and ambulant stakeholders [b,c,[Ra16]] Provides feedback function for all relevant stakeholders [a,c,[Ex09]]

[a] *Interviews;* [b] *Shadowing;* [c] *Literature*

Tab. 2: Identified use cases and responding requirements

leads to higher availability of patient data without the potential of losing data and time which may result from a manual transmission. The emphasis lies here on data security standards including the verification of sent and received documents. Patient data is highly sensitive, and in Germany, informed consent of the patient is necessary for the actors in order to be able to send those data.

3.2 Visualisation of the use cases through a process model

The "Business Process Modeling Notation" (BPMN) is suited to document existing processes, introduce new processes and to visualise changes made by digitising existing processes [FR14]. The notation is rather simple with strictly defined elements and rules that can be combined to picture complex processes. In this case, the aim is to support an existing process by implementing an IT system. Figure 1 shows the use cases stated above in BPMN.

The process shows four lanes, one for each primary actor. The lanes are clustered depending on the superordinated institution. Here, these are the hospital on the one hand and the follow-up care facility on the other. Each lane accommodates the tasks (rectangles) and events (circles) that refer to the responding primary actor. Three different lines link the separate elements. A solid line with a filled arrow defines the sequence of tasks and events, a dashed line with an empty arrow indicates the flow of messages, and a dotted line with a curved arrow links artefacts. The artefacts used in this BPMN-process are the data storage and the data object. The first exists independently from the process. For example, stored data is still available after the process of discharging a patient has ended. The latter represents information that only exists during the runtime of the process "discharge management" [FR14], i. e. data is no longer available afterwards.

The process of a patients' discharge starts with the admission of the patient to the hospital. When entering the central patient admission, the soon-to-be patient makes contact first with the nurses. They collect the patient's personal information and enter the data into the IT system. Next, they conduct a standardised assessment of the patients' need for discharge management. The result is entered into the IT system. If the patient requires discharge management, the discharge management worker is informed. If not, the process related to discharge management ends.

When a patient requires discharge management, the IT system notifies the discharge management worker. The next step is the analysis of the required proceedings and tasks. The IT system supports the discharge management worker with an overview of reported patients. It also helps with sorting out daily tasks. The discharge management worker then visits the patient in order to give information about the upcoming procedure and ask for informed consent. An assessment regarding the patient's needs and wishes is carried out. The worker directly enters all the acquired information in the IT system, and thus makes it available for following tasks and authorised staff. Especially when regarding patients in

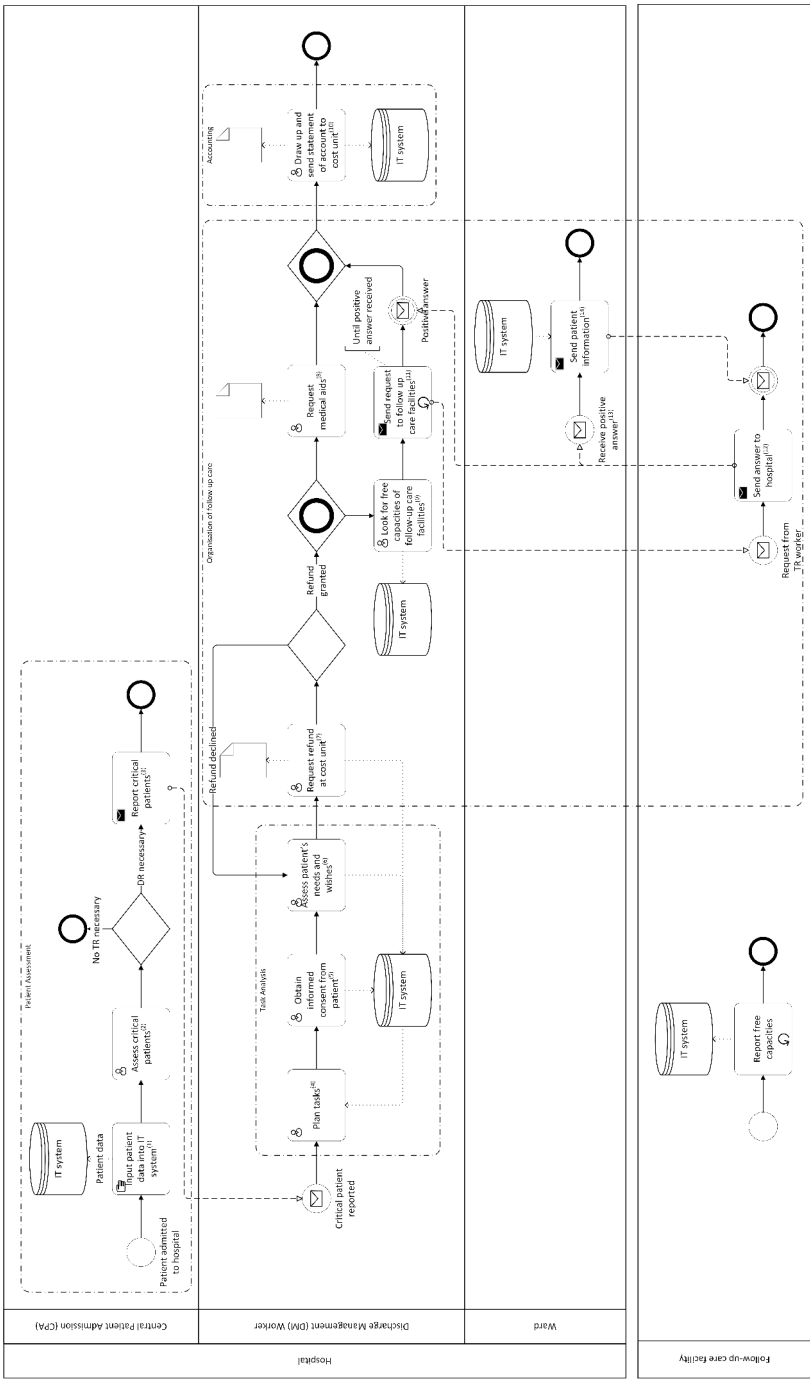


Fig. 1: Use case dependencies (BPMN Process)

need of discharge management, it is possible that they are not able to make decisions on their own. In this case, the discharge management worker consults the patient's custodian.

After the necessary steps for a gapless discharge are sorted out, the discharge management worker starts organising the follow-up care. The first step is to file applications to the responding cost units for the discussed items, such as financing nursing care or medical aids. If a refund is declined, the worker will discuss other possibilities for adequate follow-up care with the patient or custodian.

Two main tasks when organising the follow-up care are ordering needed medical aids and finding a facility that provides stationary or ambulatory care. It depends on the needs and wishes of the patient, what measurements are taken. If the patient needs medical aids, the respective form can be filled out automatically with the available patient information from the IT system. The form can also be adjusted manually and then exported or directly sent to the provider.

Since communication with follow-up care providers is necessary, the second task is more difficult. In order to circumvent manually calling a list of care facilities, the IT system shows a list with free capacities to the discharge management worker. The care facilities provide their status information at regular time intervals. The results shown can be sorted or limited to match the needs of the current patient. If a care facility is suitable, the IT system sends a request to the chosen facility. It includes essential patient information such as age, sex and necessary care. The care facility can then evaluate whether they can provide the needed care to the patient. They answer back to the discharge management worker via email. In case of a negative answer, the IT system sends a new request to another suitable care facility, is the answer positive, the ward of the patient is also notified.

The nurses at the ward maintain contact with the follow-up care facility and provide more detailed information that helps the care facility to plan for the patient's arrival — thus ensuring gapless care after discharge. The nurses are also responsible for communicating the time and date of the planned discharge as well as changes to the plan.

After the successful transition, the discharge management worker is responsible for drawing up the statement of account and sending it to the cost unit. Since all executed steps are documented in the IT system, the information can be used to fill out the predefined forms automatically. The finished forms can be exported or sent directly to the cost unit.

4 Discussion and Conclusion

In this paper the research question "What use cases can the IT-system support for a hospital discharge management and what requirements for such a system exist?" was examined.

The hospital discharge management is an interdisciplinary multi-user process. The implementation of an IT system can enhance the efficiency and effectivity of the accomplished tasks

and goals. The authors conducted multi-method research in order to analyse requirements for an IT system and derive relevant use cases. Most requirements could be identified solely by analysing the interviews and the shadowing protocol (32 of 37 requirements; 86.5%). Only eleven of the requirements resulted from both methods. The literature analysis, in combination with the analysis of the statute and the expert standard 'discharge management in healthcare', resulted in 24 out of 37 requirements (64,9%). Thus, by combining the different methods, a more holistic view of the topic could be formed.

Especially the 'expert standard discharge management' in healthcare yielded only a few requirements. The reason might become evident when regarding its aim. The expert standard provides an instrument for securing and developing quality in health care, with a focus on the improvement of the cooperation between inpatient and outpatient care. It does not constitute a set of rules for organisational processes.

The study identified mainly functional requirements. Non-functional requirements were neither named during the interviews nor found when analysing the literature. The only exception is the requirement 'portability' which originates from the shadowing. Regarding the outcome of the literature search, the lack of non-functional requirements might be due to the nature of the found literature. The publications focus mainly on the influence of new statutes regarding discharge management or on enhancing the processes from the patients' point of view. However, the development or implementation of IT systems is missing.

Nevertheless, non-functional requirements such as usability, an intuitive interface, reliability and support of learning must be taken into account, when developing IT systems. Fortunately, most of these requirements are universal and therefore already described in the scientific literature (cf. [Ba11]) as well as in DIN standards (cf. [Er06]).

The discharge of a hospital patient is a highly interdisciplinary process. For success, communication and teamwork are key aspects. The process involves not only several wards within one institution but can also require several independent ambulatory health care professionals. Being able to share the data rather than having to (re)collect, (re)input and (re)print them anew in every institution or ward, would have two main advantages. First, it would free more time resources that are very hard in need at the moment (also regarding the ongoing lack of skilled workers). Second, it would assure that up-to-date patient data is available where it is needed and when it is needed. Errors that may happen by manual transition can be diminished.

In addition to the practical benefits listed above, the development of such IT structures requires interdisciplinary research in order to understand the full range of problems and their solutions. Also, research regarding factors that ensure the participation of health care institutions on the one hand and software development enterprises on the other could be helpful. The lack of software interfaces and the associated interoperability of the systems might be one limiter regarding intersectoral communication. Therefore, we deduced the following implications for future development and research:

1. The IT-systems should support intersectoral cooperation and motivate the reduction of hierarchical structures.
2. The sharing of necessary patient data with health care professionals taking part in the patient's care should be simplified.
3. Further research should be conducted with an interdisciplinary team. The same refers to the development of software systems.

The presented research is subject to some limitations. On the one hand, the interviews and shadowing rely on one hospital with adjoining care support point in a rural area. On the other hand, although the authors interviewed various professionals in the hospital, only one discharge management worker was observed in the course of the shadowing. Even though the discharge management worker occupies the central position regarding the analysed processes, the other actors should be taken into account as well.

In future research, additional stakeholders such as physicians and pharmacists should be included. Table 3 provides an overview of the current research agenda. Each step includes a reevaluation of the previous findings. The aim is to provide a holistic approach for IT-supported discharge management, comprised of an intuitive IT system that supports the interdisciplinary processes of the discharge management and guidelines for the transfer to the IT systems of follow up healthcare institutions.

		Aim	Methods	Findings
1	Prototyping	Development of a prototype based on the requirements	Storyboards Click-Prototypes Formative Analysis	IT system in prototypic stage that can be used for real world test
2	Implementation and Validation	Evaluation of the prototype	Cognitive Walkthrough Usability Study Summative Analysis	Summative evaluation of the developed prototype
3	Transfer	Concept for transfer to other healthcare institutions	Cost-benefit analysis Quantitative study	Guideline for generalisation of IT system to other institutions

Tab. 3: Research agenda

The implications for practice and research highlight the importance of IT-systems that can support and motivate skilled workers towards a more interdisciplinary and intersectoral approach. The use cases and requirements presented in this paper shall constitute a basis and guidance when developing or improving IT-systems for discharge management but might also provide useful insights regarding software development for interdisciplinary teams in general.

References

[Ba11] Balzert, H.: Lehrbuch der Softwaretechnik: Entwurf, Implementierung, Installation und Betrieb. Spektrum Akademischer Verlag, 2011.

[Br13] Braun, J.: Entlassmanagement im Krankenhaus durch externe Leistungserbringer – (Un-)Vereinbarkeiten mit der Wahlfreiheit des Patienten. Medizinrecht 31/6, pp. 350–353, 2013.

[Br15] Braun, J.: Neue Kooperationsmöglichkeiten für Apotheken beim Entlassmanagement – zugleich eine Anmerkung zu BGH, Urteil vom 13.3.2014 – I ZR 120/13 –, MedR 2015, 36 (in diesem Heft). Medizinrecht 33/1, pp. 22–25, 2015.

[Co08] Cockburn, Alistair: Writing effective use cases. Pearson Education, Inc., 2008.

[De73] Denzin, N. K.: The Research Act: A Theoretical Introduction to Sociological Methods. Transaction Publishers, 1973.

[DM12] Deimel, D.; Müller, M.-L.: Entlassmanagement: Vernetztes Handeln durch Patientenkoordination. Georg Thieme Verlag, 2012.

[Er06] Ergonomie der Mensch-System-Interaktion - Teil 110: Grundsätze der Dialoggestaltung (ISO 9241-110:2006-08); Deutsches Institut für Normung, 2006.

[Ex09] Expertenstandard Entlassungsmanagement in der Pflege. Deutsches Netzwerk für Qualitätsentwicklung in der Pflege, 2009.

[FR14] Freund, J.; Rücker, B.: Praxishandbuch BPMN 2.0. Carl Hanser Verlag GmbH Co. KG, 2014.

[Hä12] Häser, I.: Beachtung von Apothekenrecht bei Entlassmanagement-Konzepten – Vorgaben durch das GKV-Versorgungsstrukturgesetz. Klinikarzt 41/06/07, pp. 270–272, 2012.

[Hä15] Häser, I.: Neuregelungen zum Entlassmanagement der Krankenhäuser – Ausweitung der Regelungen durch das Versorgungsstärkungsgesetz. Klinikarzt 44/12, pp. 584–585, 2015.

[Hä17] Häser, I.: Entlassmanagement – DKG klagt gegen Schiedsspruch. Klinikarzt 46/01/02, pp. 10–12, 2017.

[HK13] Hesse, S.; Klewer, J.: Anforderungen an das pflegerische Entlassungsmanagement eines Krankenhauses der Regelversorgung aus der Sicht nachsorgender Einrichtungen. HeilberufeScience 4/4, pp. 153–156, 2013.

[Hü15] Hübner, U.; Schulte, G.; Sellemann, B.; Quade, M.; Rottmann, T.; Fenske, M.; Egbert, N.; Kuhlisch, R.; Rienhoff, O.: Evaluating a Proof-of-Concept Approach of the German Health Telematics Infrastructure in the Context of Discharge Management. MedInfo 216/-, pp. 492–496, 2015.

[Ja17] Jacobs, K.; Kuhlmey, A.; Greß, S.; Klauber, J.: Pflege-Report 2017: Schwerpunkt: Die Versorgung der Pflegebedürftigen. Schattauer Verlag, 2017.

[KS16] Kuball, L.; Sailer, R.: Mehr Kompetenzen für Krankenhausärzte – Neuerungen im Entlassmanagement für flexible Anschlussversorgung. Laryngo-Rhino-Otologie 95/10, pp. 707–708, 2016.

[Ma14] Mayring, P.: Qualitative content analysis: theoretical foundation, basic procedures and software solution. Klagenfurt, 2014.

[Me17] Meinhard-Schiebel, B.: Es braucht modernes Entlassungsmanagement zur umfassenden Betreuung und Begleitung der Menschen im häuslichen Bereich. Pflege 30/4, pp. 245–246, 2017.

[MS14] Mille, M.; Stier, A.: Entlassungs- und Überleitungsmanagement. Aktuelle Urologie 45/05, pp. 381–397, 2014.

[Os15] Ossege, M.: Zum Verhältnis zwischen dem Absprachеverbot nach §11 Abs. 1 S. 1 ApoG und dem Entlassmanagement nach §39 Abs. 1 S. 4 bis 6 SGB V. Medizinrecht 33/1, pp. 36–38, 2015.

[PG12] Porter, M. E.; Guth, C.: Chancen für das deutsche Gesundheitssystem. Springer Verlag, 2012.

[PK11] Pieper, C.; Kolankowska, I.: Health care transition in Germany – Standardization of procedures and improvement actions. Journal of Multidisciplinary Healthcare 4/-, pp. 215–221, 2011.

[Ra16] Rahmenvertrag über ein Entlassmanagement beim Übergang in die Versorgung nach Krankenhausbehandlung nach § 39 Abs. 1a S. 9 SGB V. GKV Spitzenverband, 2016.

[Ro13] Romagnoli, K. M.; Handler, S. M.; Ligons, F. M.; Hochheiser, H.: Home-care nurses' perceptions of unmet information needs and communication difficulties of older patients in the immediate post-hospital discharge period. BMJ Quality & Safety 22/4, pp. 324–332, 2013.

[So05] Sommerville, I.: Integrated requirements engineering: A tutorial. IEEE Software 22/1, pp. 16–23, 2005.

[WH16] Wong, C.; Hogan, D. B.: Care Transitions: Using Narratives to Assess Continuity of Care Provided to Older Patients after Hospital Discharge. Canadian Geriatrics Journal 19/3, pp. 97–102, 2016.

[WW02] Webster, J.; Watson, R. T.: Analyzing the past to prepare for the future: Writing a Literature Review. MIS quarterly 26/2, pp. 13–23, 2002.

Digitale Transformation defizitärer Krankenhäuser in regionale Pflegekompetenzzentren

Christian Fitte[1] und Frank Teuteberg[1]

Abstract: Während der Bedarf an Pflegeplätzen für ältere Menschen unaufhaltsam wächst, verzeichnen Krankenhäuser, insbesondere in ländlichen Regionen, oftmals finanzielle Verluste aufgrund leerstehender Betten und geringer Fallzahlen. Daher wird in diesem Beitrag ein Konzept vorgestellt, mit dem defizitäre Krankenhäuser in Pflegekompetenzzentren umgewandelt werden können. Zentraler Bestandteil ist eine auf Informations- und Kommunikationstechnologie (IKT) basierende Infrastruktur, die alle Akteure besser miteinander vernetzt, die Arbeit für Pflegekräfte, Ärzte und Apotheken vereinfacht und Angebote für Patienten besser zugänglich macht. Im Rahmen eines Design Science Research-Ansatzes werden Stakeholder und Anforderungen an ein IKT-basiertes Pflegekompetenzzentrum erhoben. Nach einer Konzeptbeschreibung des regionalen Pflegekompetenzzentrums (Reko) und einer Diskussion werden Evaluationsmöglichkeiten aufgezeigt.

Keywords: Pflege, eHealth, defizitäre Krankenhäuser, Digitale Patientenakte, regionales Pflegekompetenzzentrum

1 Einleitung und Motivation

Aufgrund des demografischen Wandels steigt der Anteil älterer Menschen an der Gesamtbevölkerung in Deutschland kontinuierlich an. Nach aktuellen Prognosen wächst der Anteil der über 80-Jährigen bis 2050 um 156 % [Pe19]. Damit erhöht sich auch der Bedarf für ambulante und stationäre Pflege. Obwohl die Anzahl an Pflegeheimen steigt, besteht in vielen Regionen bereits heute ein Mangel an ausreichenden Versorgungsplätzen für ältere Menschen. In der Folge kommt es zu einer Überlastung der Pflegekräfte und zu Qualitätsverlusten in der Pflege [Ge17]. In ländlichen Regionen ist dieser Effekt verstärkt zu beobachten, da dort vermehrt ältere Menschen leben und insbesondere die ambulante Pflege aufgrund langer Distanzen mehr Zeit in Anspruch nimmt. Viele junge Menschen ziehen wegen besserer Berufsperspektiven in umliegende Städte und stehen nicht mehr als Arbeitskräfte in der Pflege zur Verfügung.
Während in der Pflegeversorgung auf dem Land große Engpässe existieren, bestehen gleichzeitig häufig Überkapazitäten in ländlichen Krankenhäusern. Wegen der sinkenden Bevölkerungszahl haben ländliche Krankenhäuser geringe Fallzahlen und zunehmend ungenutzte Kapazitäten [BDO14]. Die vorhandenen Betten werden dann meist von älte-

[1] Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49076 Osnabrück, {christian.fitte; frank.teuteberg}@uni-osnabrueck.de

doi:10.18420/inf2019_10

ren Menschen belegt: Etwa die Hälfte der Patienten[2] in Allgemeinkrankenhäusern ist älter als 65 Jahre. Die Anzahl von Geriatriepatienten hat sich von 2006 bis 2015 verdreifacht und etwa 12 % sind von einer Demenzerkrankung betroffen [AHP17], [DAG18]. Diese Umstände führen zu zwei erheblichen Nachteilen: Einerseits sind ländliche Krankenhäuser vermehrt unwirtschaftlich und verzeichnen Verluste, andererseits sind diese defizitären Krankenhäuser meist nicht auf die Pflege älterer Menschen spezialisiert. Durch die unsachgemäße Versorgung sinkt das Wohlbefinden der Patienten. Die ungewohnte Umgebung und der Mangel an speziell geschultem Pflegepersonal führen bei älteren Menschen nicht selten zu einem sog. Krankenhausdelirium. Dabei werden physische Fähigkeiten der Patienten schneller abgebaut, sie liegen im Schnitt vier Tage länger im Krankenhaus, haben mehr Nebendiagnosen und ein sieben Mal höheres Risiko dort zu versterben [Gu17].

Um einerseits der Überversorgung von Krankenhauskapazitäten und andererseits der Unterversorgung von Pflegeplätzen in ländlichen Regionen entgegenzuwirken, wird in dem Forschungsprojekt „Regionales Pflegekompetenzzentrum – Innovationsstrategie für die Langzeitversorgung vor Ort“ die Umwandlung eines ländlichen, von Schließung bedrohtem Krankenhauses in ein regionales Pflegekompetenzzentrum (Reko) angestrebt [KM18]. Das Konzept wurde von der DAK-Gesundheit (Konsortialführer) vorgelegt und wird in einem vierjährigem, vom Innovationsfond mit ca. 10 Millionen Euro gefördertem Forschungsprojekt in der Modellregion Landkreis Grafschaft Bentheim/Landkreis Emsland mit der Gesundheitsregion EUREGIO e. V. umgesetzt. Ziel ist es, vorhandene Krankenhausinfrastrukturen für die stationäre Pflege zu nutzen und alle weiteren Akteure für die ambulante Versorgung zusammenzuführen. Gleichzeitig sollen ausreichend Kapazitäten für eine Grundgesundheitsversorgung der Bevölkerung erhalten bleiben. Wesentliches Merkmal ist, dass durch eine sektorenübergreifende Informations- und Kommunikationstechnologie (IKT) die Akteure nicht nur physisch, sondern auch informationstechnisch besser miteinander vernetzt werden. Der Einsatz soziotechnischer Systeme kann den Arbeitsalltag der Pflegekräfte, pflegender Angehöriger und das Leben der Pflegebedürftigen erleichtern [Mü12].

Der vorliegende Beitrag ist wie folgt aufgebaut: Zunächst werden in Abschnitt zwei Grundlagen der Gesundheitsversorgung in ländlichen Regionen, verwandte Projekte sowie eine daraus identifizierte Forschungslücke herausgearbeitet. Abschnitt drei beschreibt den methodischen Rahmen der Konzeption, welcher dem Design Science Research-Ansatz folgt. Abschnitt vier beinhaltet eine Anforderungsanalyse, eine Beschreibung der IKT-Infrastruktur sowie das Zielartefakt. Nach einer Diskussion der Ergebnisse werden Evaluationsmöglichkeiten aufgezeigt. Es folgt ein zusammenfassendes Fazit sowie ein Ausblick auf die zukünftige Projektarbeit.

[2] Aus Gründen der besseren Lesbarkeit wird auf die Nennung der weiblichen Sprachform verzichtet. Sämtliche Personenbezeichnungen gelten gleichermaßen für beide Geschlechter.

2 Gesundheitsversorgung in ländlichen Regionen

2.1 Fehlallokation vorhandener Ressourcen

In Deutschland gibt es circa 2.000 Krankenhäuser, von denen sich rund 600 in ländlichen, strukturschwachen Regionen befinden [BDO14]. Diese Krankenhäuser sind im Gegensatz zu städtischen Einrichtungen häufig die einzige Anlaufstelle für eine wohnortnahe Gesundheitsversorgung und stellen gleichzeitig einen bedeutenden Arbeitgeber in der Region dar. Dennoch stehen ländliche Krankenhäuser vor besonderen Herausforderungen: Aufgrund von mangelnden Spezialisierungs- und Kooperationsmöglichkeiten, sowie geringer Fallzahlen im Einzugsgebiet bleiben zahlreiche Betten unbelegt [AHP17], [BDO14]. In der Folge verzeichnen ländliche Krankenhäuser vermehrt finanzielle Verluste, sodass viele Einrichtungen von einer Schließung bedroht sind [BDO14]. Besonders kleine Einrichtungen mit weniger als 200 Betten werden voraussichtlich nicht mehr lange bestehen können oder müssen erhebliche Umstrukturierungen vornehmen. Zu diesem Zweck wurde der Krankenhausstrukturfond eingerichtet, der vorsieht, Krankenhauskapazitäten abzubauen bzw. umzuwandeln [KL15]. Patienten müssten folglich lange Wege in Nachbarorte zurücklegen, zahlreiche Arbeitsplätze würden wegfallen und die Attraktivität der Region würde sinken. Insbesondere vor dem Hintergrund eines Landärztemangels ist der Erhalt einer Grundversorgung durch ländliche Krankenhäuser von großer Bedeutung. Gleichzeitig müssen Lösungen für den wachsenden Bedarf an Pflegeplätzen entwickelt werden.

2.2 Verwandte Projekte

Der Einsatz von IKT zur Verbesserung der ambulanten und stationären Pflege ist bereits Bestandteil zahlreicher Projekte in der Forschung und in der Praxis, welche anhand einer Onlinerecherche identifiziert wurden. Projekte bei denen digitale Technologien zur Verbesserung der Pflegeversorgung eingesetzt wurden sind in Tabelle 1 zusammengefasst.

Name	Beschreibung	Literatur
Pflegeinnovationszentrum (PIZ)	Der Einsatz neuer Technologien wird in den vier Einsatzszenarien häusliche Pflege, stationäre Pflege, Intensivpflege und in der Pflegedienstzentrale untersucht.	[BHH18]
Dorfgemeinschaft 2.0	Entwicklung digitaler Lösungsmöglichkeiten in den vier Lebenswelten Wohnen, Versorgung, Mobilität sowie Gesundheit und Pflege. Alle Angebote können über den „digitalen Dorfmarktplatz“ gebucht werden.	[FTI16]
Gestaltung altersgerechter Lebenswelten (GAL)	Interdisziplinäres Forschungsprojekt, bei dem der Einsatz von Ambient Assisted Living (AAL) Technologien in den Szenarien Haushaltsassistenz, Prävention und Monitoring im Rehasport, sensorgestützte Aktivitätsbestimmung und Sturzprävention untersucht wurde.	[Ha14]
ITAGAP	Entwicklung technikgestützter Systeme, mit denen Arbeitsprozesse, wie z. B. die Pflegeplanung und -dokumentation unterstützt werden. ITAGAP bedeutet Integrierte Technik- und Arbeitsprozessentwicklung	[Br17]

Name	Beschreibung	Literatur
	für Gesundheit in der ambulanten Pflege.	
ATMoS-PHÄRE	Partnern aus Wissenschaft und Praxis untersuchen, wie eHealth-Innovationen die Lebensqualität von multimorbiden Patienten erhöhen.	[At18]
Netzwerk GesundAktiv	Unterstützung der häuslichen Pflege durch den Technikassistenten PAUL (Persönlicher Assistent für unterstütztes Leben). Unterhaltungs- und Gesundheitsangebote können über ein Tablet genutzt werden.	[Sp10]
senimed-IT	In dem Projekt wird die intersektorale Vernetzung von Hausärzten und Pflegeeinrichtungen gefördert, um die Patientenversorgung in der stationären Pflege zu verbessern [Se18]. Wesentliche Bestandteile sind eine gemeinschaftliche Dokumentation, ein vernetzter Medikationsplan sowie ein Frühwarnsystem.	[Se18]
eHome	Überstützung der Zusammenarbeit in der Pflege durch eine Plattform. Ambulante Pflegekräfte können Symptome oder Probleme mit bestimmten Medikamenten online vermerken und direkt an einen Arzt oder Apotheker weiterleiten.	[Di18]
Optimierte Arzneimittel-versorgung (OAV)	Projekt zur Förderung der sektorenübergreifenden Kommunikation. Die verbesserte Kommunikation von Ärzten, Apothekern, Pflegeeinrichtungen und Patienten soll unerwünschte Arzneimittelereignisse (UAE) vermeiden.	[HHF18]

Tab. 1: Übersicht verwandter Forschungsprojekte

Die vorgestellten Projekte zeigen die Potenziale der Digitalisierung in unterschiedlichen Anwendungsfeldern der Pflege auf und kommen zu dem Ergebnis, dass eine sektorenübergreifende Vernetzung der einzelnen Akteure in der Pflege von besonderer Bedeutung sei. Ansätze für eine Vereinheitlichung der Informationsinfrastruktur über alle Akteure existieren jedoch nicht, woraus sich die Forschungslücke ergibt, welche mit dem vorliegenden Beitrag geschlossen wird. Das in diesem Beitrag vorgeschlagene Konzept unterscheidet sich von bisherigen Lösungen dahingehend, dass es einen holistischen Ansatz verfolgt, bei dem alle an der Gesundheits- und Pflegeversorgung beteiligten Akteure physisch und informationstechnisch miteinander vernetzt werden. Hierfür werden vorhandene Krankenhausinfrastrukturen, die bislang nicht effizient genutzt werden, zu Pflegekompetenzzentren umfunktioniert. Eine für alle beteiligten Akteure zugängliche Plattform unterstützt die sektorenübergreifende Zusammenarbeit. Zu dessen Konzeption besteht weiterer Forschungsbedarf. Für den Aufbau des Rekos ergeben sich somit folgende Forschungsfragen (FF):

FF1: Welche Stakeholder sind bei der Konzeption eines regionalen Pflegekompetenzzentrums zu berücksichtigen?

FF2: Welche speziellen Anforderungen haben die ermittelten Stakeholder an die IKT-Infrastruktur eines regionalen Pflegekompetenzzentrums?

FF3: Wie können die Anforderungen aller Stakeholder als Gestaltungselemente einer IKT-Infrastruktur umgesetzt werden?

Die Ergebnisse aus den verwandten Projekten fließen bei der Anforderungserhebung an das regionale Pflegekompetenzzentrums gemäß Tabelle 3 ein.

3 Forschungsmethode

Zur Beantwortung der ermittelten Forschungsfragen wird die Design Science Research Methode (DSR) genutzt. DSR ist als Forschungsansatz definiert, in dem durch die Entwicklung innovativer Konzepte gesellschaftlich relevante Probleme gelöst werden und gleichzeitig ein Zugewinn für die Wissenschaft generiert wird [HC10]. Einerseits werden mit dem Konzept des Rekos akute Probleme in der Pflegeversorgung adressiert, andererseits wird durch die systematische Aufarbeitung der Anforderungen und Konzeption neues Wissen generiert. Die Konzeption einer Lösungsstrategie wird von den Faktoren Umgebung und Wissensbasis beeinflusst, wie Abbildung 1 veranschaulicht [He04].

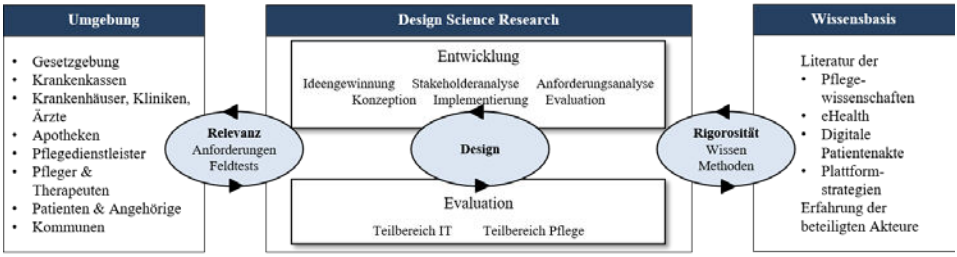


Abb. 1: Design Science Ansatz bei der Reko Konzeption (Hevner et al. 2007)

Die Umgebung beschreibt den Problembereich in dem sich das behandelte Thema befindet. Dabei spielen alle betroffenen Akteure, institutionelle und rechtliche Rahmenbedingungen sowie die relevanten Technologien eine bedeutende Rolle. Die Verbindung zwischen der Umgebung und dem Kern des DSR besteht durch die **Relevanz** [He04]. Die Wissensbasis repräsentiert bisherige Erkenntnisse aus Disziplinen, die mit dem betrachteten Problemfeld verwandt sind. Die **Rigorosität** stellt sicher, dass die erstellte Lösung im Einklang mit bisherigen wissenschaftlichen Erkenntnissen steht und dass das Zielartefakt methodisch nach wissenschaftlichen Standards erarbeitet wird. Im Kern des DSR steht der **Design Zyklus**, welcher aus einem iterativem Prozess der Entwicklung neuer Ergebnisse und deren Evaluation besteht.

Nach der Ideengewinnung wurde zunächst eine Übersicht der beteiligten Stakeholder erarbeitet. Durch eine Literaturanalyse nach Webster und Watson (2002) mit den Stichworten `Pflege OR Care AND IKT OR ICT OR Technolog* OR Digital* OR Plattform OR platform` in sechs wirtschaftsinformatikspezifischen Datenbanken wurden 15 verwandte Forschungsarbeiten und Projekte identifiziert [WW02]. Auf dieser Basis sowie einer qualitativen Querschnittsanalyse in Form von fünf Workshops mit je vier bis acht Teilnehmern und zehn semistrukturierten Experteninterviews [MN07] wurden anschließend die Akteure und deren jeweiligen Anforderungen an das Reko erhoben. So wurden im Zeitraum von November 2017 bis März 2018 insgesamt 38 Akteure aus dem Gesundheitswesen sowie aus den Forschungsbereichen Pflegewissenschaften und eHealth konsultiert. Die Ergebnisse wurden protokolliert

und von den Autoren unabhängig voneinander ausgewertet. Die konsolidierte Anforderungsanalyse ist Bestandteil von Abschnitt 4.1. In Abschnitt 4.2 wird das digitale Ökosystem mit seinen Bestandteilen vorgestellt. Zielartefakt ist das Modell des Rekos, welches in Abschnitt 4.3 vorgestellt wird. Im Rahmen der Diskussion in Abschnitt 5 wird die Evaluation der Ergebnisse thematisiert. Bei der Durchführung des DSR-Ansatzes wurden die Richtlinien gemäß Tabelle 2 beachtet [He04].

Nr.	Richtlinie	Umsetzung
1	Artefakt	Zielartefakt ist das in Abschnitt 4.3 vorgestellte Konzept.
2	Relevanz	Mit dem wachsenden Bedarf an Pflegekapazitäten adressiert das Reko ein praktisches und gesellschaftlich relevantes Problem.
3	Evaluation	Nutzeneffekte, Kosten und Wirkungsmechanismen werden durch eine randomisierte Kontrollstudie mit zwei Vergleichsgruppen überprüft.
4	Forschungsbeitrag	Forschungsbeitrag zur Konzeption und Evaluation von Umstrukturierungen defizitärer Krankenhäuser.
5	Rigorosität	Durch strukturierte Literaturrecherchen [WW02], Experteninterviews [MN07] und Workshops wird eine wissenschaftliche Vorgehensweise verfolgt.
6	Design als Prozess	Durch kontinuierliche Abstimmungen mit beteiligten Akteuren erfüllt das Reko die rechtlichen und stakeholderspezifischen Anforderungen.
7	Ergebnisveröffentlichung	Ergebnisse des Rekos werden auf wissenschaftlichen Tagungen präsentiert, der Fachpresse vorgestellt und über lokalen Medien der Öffentlichkeit zugänglich gemacht.

Tab. 2: Umsetzung der DSR Richtlinien (Hevner et al. 2004)

4 Konzeption eines regionalen Pflegekompetenzzentrums

4.1 Anforderungsanalyse

Durch Workshops mit unterschiedlichen Vertretern der Gesundheitsbranche (Krankenkasse, Gesundheitsnetzwerk, Wissenschaftlern aus dem Bereich eHealth) wurden zunächst folgende Stakeholder identifiziert, die im Rahmen des Rekos berücksichtigt werden: Pflegedienstleister, Kliniken und Ärzte, Apotheken, Pfleger und Therapeuten, Patienten und Angehörige, Krankenkassen sowie die Kommunen. Anschließend wurden anhand von Literatur- und Projekt-Recherchen, Experteninterviews [E] und Workshops [W] die Anforderungen der einzelnen Stakeholder an das Reko gemäß Tabelle 3 ermittelt.

Für **Pflegedienstleister** sollen Prozesse und Strukturen innerhalb des Rekos effizienter gestaltet werden [Br17, E, W]. Alle Schnittstellen werden in einem Gebäude zusammengeführt, wodurch kürzere Kommunikationswege und Bearbeitungszeiten erzielt werden. Neben einem Bereich, in dem Krankenhauskapazitäten untergebracht werden, soll im Reko auch ein Hausarzt vorhanden sein [E]. So können **Kliniken und Ärzte** ein verbessertes Entlassmanagement realisieren, und Patienten müssen keine langen Transportwege zurücklegen [E]. Der Übergang aus einer Krankenhausbehandlung in die Pflege ver-

läuft fließend und wichtige Behandlungsdaten werden den Pflegefachkräften über eine elektronische Patientenakte zugänglich gemacht [Se18, E, W]. Bei etwaigen Rückfragen seitens des Pflegepersonals sind die Ärzte direkt vor Ort [Se18, E]. Für ambulant pflegebedürftige Patienten soll das ärztliche Personal im Reko via Telemedizin Fragen beantworten und den Gesundheitszustand der Patienten überprüfen können [Sp10, Br17, E]. Mit steigendem Alter erhöht sich häufig auch die Anzahl an einzunehmenden Medikamenten, womit das Risiko UAE steigt. Um diese zu vermeiden werden **Apotheken** im Reko integriert, um regelmäßige Medikationschecks durchzuführen [E]. Mit einer elektronischen Medikationsübersicht [Sp10, E] können Wechselwirkungen unterschiedlicher Präparate automatisiert aufgedeckt werden, wodurch die Arzneimitteltherapiesicherheit bei Polymedikation steigt [RK15]. Neben Telemedizin bestehen auch Angebote zur Telepharmazie [W]. Ein weiterer Anwendungsfall zur besseren Vernetzung der Akteure ist das eRezept [Di18].

Besonders wichtige Akteure im Reko sind **Pfleger und Therapeuten**. Für sie soll im Rahmen des Projektes eine spürbare Arbeitsentlastung erzielt werden, in dem Verwaltungsaufgaben über die Plattform vereinfacht werden, sodass mehr Zeit für die Pflege der Patienten bleibt [Br17, E, W]. Neuartige Technologien zur Unterstützung der Pflege, wie z. B. Augmented Reality (AR) oder Pflegeroboter sollen kontinuierlich getestet werden [BHH18, At18, An17, JP16, JP17, E].

Ebenfalls zentrale Akteure sind die pflegebedürftigen **Patienten und deren Angehörige**. Patienten in der stationären Pflege sollen von dem Reko profitieren, indem sie für Arztbesuche keine beschwerlichen Wege zurücklegen müssen [E], sondern der Arzt zu den Patienten kommen kann. Außerdem können weitere Dienstleistungen, wie z. B. ein Mobilitätsservice [FTI16, W], Kultur- und Veranstaltungsprogramme [W], Beratungen [KM18, E, W] u. v. m. über eine zentrale Plattform abgerufen werden. Im Bereich der ambulanten Pflege stehen Telemedizinische und -pharmazeutische Beratungsmöglichkeiten zur Verfügung [Sp10, Br17, E]. Auch Mobilitäts- und Versorgungsservices, wie Essen auf Rädern oder ein Medikamentenlieferdienst, können genutzt werden [W]. Ein weiterer Bestandteil der ambulanten Pflege sind AAL- und Smart Home-Technologien, welche über die zentrale Plattform verwaltet werden können [Ha14, Sp10, Di18, Be17, W]. Hierzu zählen Sensoren zur Überwachung der Vitalparameter und Notfallsysteme z. B. im Falle eines Sturzes [Be17].

Angehörige der Pflegebedürftigen können ebenfalls Dienste über die Plattform buchen und verwalten. Für sie ist die Beratungsfunktion von großer Bedeutung [KM18, W]. Häufige Fragen sollen bereits über die Plattform beantwortet werden [E], weitergehende Beratungen können schließlich in den Räumlichkeiten des Rekos durchgeführt werden [E, W]. Bei Patienten und Angehörige ist in besonderer Weise auf eine benutzerfreundliche, altersunabhängige und intuitive Bedienung zu achten [Mü12, E, W]. Die Benutzer sollen mit der Technik keinesfalls überfordert werden [W]. Aus diesem Grund sollte diese Gruppe bereits bei der Gestaltung der Services mit einbezogen werden [Pr12].

		Akteure						Quelle				
		Pflegedienstleister	Kliniken & Ärzte	Apotheken	Pfleger & Therapeuten	Patienten & Angehörige	Krankenkassen	Kommunen	Literatur	Verwandte Projekte	Experteninterviews	Workshops
Funktionale Anforderungen												
Telemedizin	Echtzeitkommunikation		x	x	x	x			[Br17]			x
	Digitale Sprechstunde		x	x	x	x			[Br17]	[Sp10]	x	
	Telepharmazie			x	x	x			[Br17]			x
	(Video-) Chat-Funktion		x	x	x	x				[Sp10]		x
Intersektorale Zusammenarbeit	Entlassmanagement	x	x		x	x				[Br17]	x	x
	Intersektoraler Dokumentenaustausch	x	x	x	x	x	x				x	
	Digitaler Medikationsplan		x	x	x	x				[Sp10,Se18]	x	
	Digitale Patientenakte		x	x	x	x	x			[Se18]	x	x
	Elektronisches Rezept		x	x		x	x		[Di18]		x	
	Bestellmanagement	x	x	x	x	x	x					x
	Entscheidungsunterstützungssysteme	x	x	x	x	x	x	x			x	
Patientenservices	Mobilitätsservices	x				x		x	[KM18]	[FTI16]		x
	Medikamentenlieferdienst			x		x						x
	Essen auf Rädern	x				x			[KM18]			x
	Kultur- und Unterhaltungsprogramm	x				x		x	[KM18]			x
	Beratungsservice	x	x	x	x	x	x	x	[KM18,Be17]	[Sp10]	x	x
	Motivation für gesundheitsfördernde Maßnahmen		x	x		x					x	
Ambulante Pflege	Interaktive Assistenzfunktion	x	x		x	x						x
	Smart Home/ AAL	x			x	x			[Di18,Be17]	[Ha14,Sp10]		x
	Hausnotrufsystem	x			x	x			[Pr12]	[Ha14]		x
	Sensormatten	x			x	x			[Di18]			
	Medikamentenmonitoring		x	x	x	x			[RK15]		x	
	Erinnerungsfunktion			x		x					x	
Verwaltung	Verwaltungsangelegenheiten	x			x		x	x		[Br17]	x	x
	Abrechnungen	x	x	x		x	x			[Br17]		x
	öffentliche Verwaltung	x				x		x			x	x
	Intelligente Berichterstellung	x	x									
Nicht-Funktionale Anforderungen												
Benutzung	Geräteunabhängigkeit	x	x	x	x	x	x	x			x	
	Einbindung neuer Technologien (z.B. AR/VR)	x	x	x	x	x			[An17,JP16, JP17]	[BHH18, At18]	x	
	Benutzerfreundlichkeit				x	x			[Mü12,Pr12]		x	x
	Benutzerspezifische Oberfläche	x	x	x	x	x	x	x	[Mü12]		x	x
	Anleitungen und Hilfestellungen	x	x	x	x	x					x	
Sicherheit	Schutz der Privatsphäre		x	x		x					x	
	Datenschutz	x	x	x	x	x	x	x			x	x
	Datensicherheit	x	x	x	x	x	x	x			x	x
	Einhaltung von IETF-, W3C, ISO und IEC Standards	x	x	x							x	

Tab. 3: Anforderungen der Stakeholder an das regionale Pflegekompetenzzentrum

Als zentraler Ansprechpartner für die Abrechnung von Gesundheitsleistungen werden auch **Krankenkassen** in das digitale Ökosystem des Rekos eingebunden, um bei Abrechnungsprozessen die Kommunikation zu vereinfachen [Br17, W]. Abschließend sollen auch die **Kommunen** bei der Konzeption des Rekos berücksichtigt werden [E, W]. Häufig müssen Pfleger oder Angehörige Behördengänge erledigen und haben somit weniger Zeit für die Betreuung [E]. Durch die Einbindung der Kommunen werden Prozesse vereinfacht und die Kommunikation erleichtert.

Bei der Umsetzung der identifizierten Anforderungen ist in besonderem Maße auf eine zielgruppengerechte Ansprache zu achten [E]. Für das Pflegepersonal sollen die Anwendungen eine spürbare Arbeitserleichterung darstellen [Br17], [Ge17]. Für Pflegebedürftige und deren Angehörige sollen alle Angebote möglichst einfach zugänglich gemacht werden. Hierbei sind bisherige Erkenntnisse aus dem Design von IKT-Anwendungen für ältere Menschen im Pflegebereich zu beachten [Mü12]. Die Anforderungen der jeweiligen Akteure sowie deren Quellen sind in Tabelle 3 in den Kategorien funktionale und nicht-funktionale Anforderungen zusammengefasst.

4.2 IKT-Infrastruktur

Auf Basis der identifizierten Anforderungen wird im Rahmen des Rekos ein digitales Ökosystem zum besseren Informationsaustausch bei den Sektorenübergängen erstellt. Das digitale Ökosystem ermöglicht einen zentralen Datenzugriff und eine transparente Darstellung von Gesundheitsdaten sowie ein selbstbestimmtes Teilen durch die Patienten. Somit ergeben sich u. a. folgende Vorteile: Kontinuierlicher Einblick in medizinische Daten, Mehrwert durch personalisierte Angebote sowie eine vereinfachte Kommunikation und sektorenübergreifender Datenaustausch. Für Patienten, die Hilfe beim Management ihrer Daten benötigen, ist ein Zugriff Dritter (z. B. Case Manager, Familienangehörige) möglich. Zur Veranschaulichung der IKT-Infrastruktur wurde eine Architektur des digitalen Ökosystems gemäß Abbildung 2 mit fünf Schichten konzipiert.

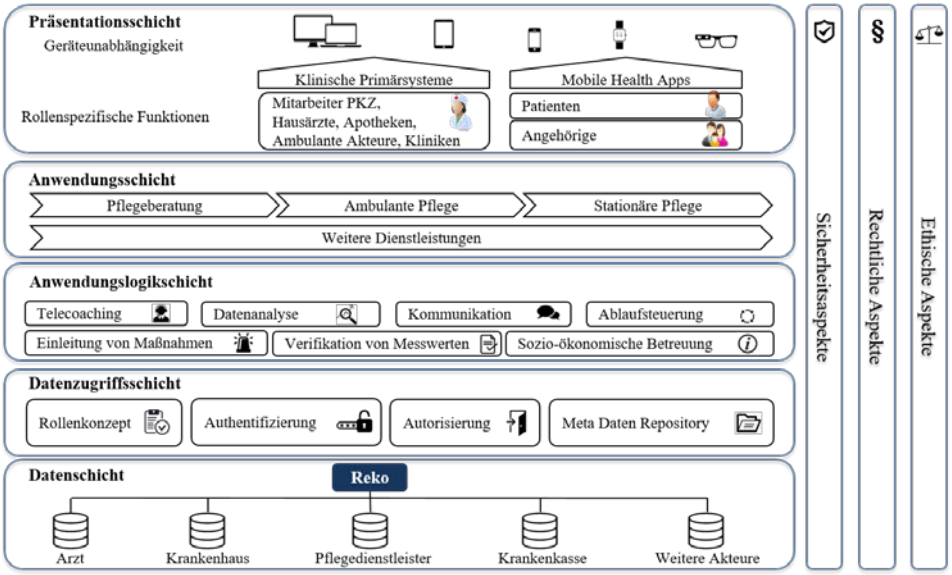


Abb. 2: IKT-Architektur und beispielhafte Anwendungen des Rekos

Im digitalen Ökosystem stehen auf dem FHIR-Standard (Fast Healthcare Interoperability Resources), den Standard-Profilen der IHE (Integrating the Healthcare Enterprise), der HL7-Familie (Health Level 7) sowie der xDT-Familie definierte Kommunikationsmuster und Schnittstellen zum strukturierten Informations- und Dokumentenaustausch bereit. Eine Konformität mit der ISO 21090 (Health informatics - Harmonized data types for information interchange) und der ISO 13606-Familie (Health informatics - Electronic health record communication) wird gewährleistet.

Die Module des digitalen Ökosystems lassen sich in die Kategorien Verwaltung, Pflege und Gesundheitsversorgung und Services für Patienten und Angehörige gliedern. Im Bereich Verwaltung können Anwendungen zum Personalmanagement, zum Zugriffsmanagement, zur Abrechnung, zum Versorgungsmanagement, Applikationen für Fortbildungsmaßnahmen sowie Catering Verwaltungsfunktionen genutzt werden. Im Bereich Pflege und Gesundheitsversorgung können über das digitale Ökosystem eine digitale Patientenakte (z. B. auf der Basis von *vivy*; www.vivy.com), sowie ein Medikations- und Therapieplan genutzt werden. Weiterhin sind Telemedizin Angebote verfügbar und Patienten können über Sensoren gemessene Vitalparameter mit ihrem Arzt teilen. Im Bereich Services für Patienten und Angehörige können über die digitale Plattform Smart Home und AAL Systeme verwaltet werden. Es kann ein Hausnotrufsystem eingerichtet werden, Fahrdienste können gebucht werden, Gesundheits- und Pflegeberatungen werden angeboten und es können aktuelle Kultur- und Unterhaltungsprogramme eingesehen werden.[3] Da die Plattform offen und flexibel gestaltet wird, ist der Modulkatalog nicht auf die genannten Bestandteile begrenzt, sondern kann bei Bedarf um weitere Anwendungen auch von Drittanbietern erweitert werden. Da auf der Plattform höchst sensible Gesundheitsdaten gespeichert werden, kommt dem Datenschutz eine besonders hohe Bedeutung zu. Hierbei gilt der Grundsatz: Der Nutzer bleibt Herr der eigenen Daten. Somit bleibt dem Nutzer selbst überlassen, welche Daten gespeichert werden, wem er wann Zugriff auf welche Daten geben möchte, und welche Daten für welche Dienste verwendet werden dürfen. Sämtliche Daten sind nur über eine Zwei-Faktor-Authentifizierung zugänglich.

4.3 Konzept zur Zusammenführung umfassender Pflegeservices

Nach der systematischen Anforderungserhebung und der Definition der IKT-Infrastruktur wurde das Konzept des Rekos erstellt. Im Kern stehen die fünf Grundbausteine Pflegeberatung, Ambulante Pflege, Stationäre Pflege, Medizinische Versorgung und Mobilitätsdienstleistungen, die im Pflegekompetenzzentrum unter einem Dach zusammengeführt werden [KM18]. Das Fundament der Zusammenarbeit bildet die technikgestützte Infrastruktur, durch die alle Akteure sowie deren Angebote und Dienste zentral zusammengeführt werden. In diesem Zusammenhang ist die digitale Patientenakte von besonderer Bedeutung, auf deren Grundlage die Akteure ihre Zusammenarbeit

[3] Eine Übersicht der Module mit einer detaillierten Beschreibung kann unter https://bit.ly/2LrCnQy oder auf www.rekopflege.de eingesehen werden.

gestalten können. Ergänzt wird das Grundangebot des Rekos durch weitere Komplementärfunktionen.[3] Im Rahmen eines Bildungszentrums wird allen professionellen Akteuren die Möglichkeit gegeben, sich in ihrem Fachgebiet weiterzuqualifizieren. Möglichkeiten zur Freizeitgestaltung sollen das Leben der pflegebedürftigen Bewohner in der Einrichtung bereichern. Neben Angeboten zur stationären Pflege wird durch ein Case Management auch die ambulante Pflege unterstützt. Durch den holistischen Ansatz des Rekos wird einerseits die Qualität der Pflege verbessert, indem UAE oder ein Krankenhausdelirium vermieden werden, andererseits werden für alle professionellen Akteure, die an der Pflege beteiligt sind, die Arbeitsbedingungen verbessert. Von dem erzielten Zeitgewinn profitieren dann ebenfalls die Patienten durch eine qualitativ bessere Versorgung.

5 Diskussion und zukünftige Evaluation

Die Transformation ländlicher, defizitärer Krankenhäuser in regionale Pflegekompetenzzentren verfolgt zwei Ziele: Einerseits werden ungenutzte Krankenhauskapazitäten sinnvoll umgewandelt, andererseits wird ein Beitrag zum steigenden Pflegebedarf geleistet. Durch die physische und informationstechnische Vernetzung von sozialen, medizinischen und pflegerischen Lösungsanbietern innerhalb des Rekos können Daten, Anwendungen und Services miteinander verknüpft werden und zu einer Kosteneinsparung im Gesundheitswesen beitragen. Zudem wird die Versorgungsqualität vom individuellen Wohnort entkoppelt. Anhand einer Analyse der Auslastungs- und Nutzungsdaten können weitere Verbesserungsvorschläge generiert werden.
Der bloße Einsatz von IKT kann jedoch nicht als Universallösung für alle Probleme in der Pflege angesehen werden. Vielmehr soll IKT dazu beitragen, dass Akteure in der Pflege entlastet werden, damit sie sich wieder auf ihre Kernaufgaben fokussieren können. Daher müssen Lösungsmöglichkeiten in enger Zusammenarbeit mit den Hauptanwendern entwickelt werden, um die gewünschten Nutzeffekte erzielen zu können. Insofern wird sich die Konzeption der o. g. Plattform nicht ausschließlich an den bereits identifizierten Anforderungen orientieren, sondern Lösungsvorschläge werden in iterativen Workshops interdisziplinären Gruppen vorgestellt und an das erhaltene Feedback angepasst. Prototypen werden schließlich in Feldtests durch systematische Befragungen, Beobachtungen (Shadowing) sowie Kosten-, Prozess- und IKT-Systemanalysen evaluiert und kontinuierlich verbessert.

Zur Evaluation der Wirksamkeit des Rekos wird eine cluster-randomisierte, longitudinale Studie durchgeführt [Ch09]. Hierfür werden die Patienten in eine Interventions- sowie Kontrollgruppe eingeteilt. Die Evaluation ist in die Bestandteile Pflege und IKT-Konzept gegliedert. Im Teilbereich IKT-Konzept wird herausgearbeitet, ob die IKT-Infrastruktur die gewünschten Nutzeffekte erzielt. Dabei sollen durch eine formative und summative Evaluation insbesondere folgende Fragestellungen beantwortet werden:

- Werden pflegende Angehörige durch die Reko-Angebote spürbar entlastet?

- Kann durch den Einsatz der Reko-Plattform sowie durch AAL-Technologien eine bessere Lebensqualität für Pflegebedürftige erzielt werden?
- Ist die Nutzung der Plattform intuitiv und kann die Zielgruppe alle Angebote ohne Schwierigkeiten nutzen?
- Werden professionelle Pflegekräfte durch das IKT-Konzept spürbar entlastet?
- Wird die intersektorale Kommunikation signifikant verbessert?
- Wird die IKT-Infrastruktur Anforderungen zur Funktionalität, Sicherheit, Interoperabilität, Zuverlässigkeit, Effizienz gemäß DIN ISO/IEC 25000 gerecht?

Im Teilbereich Pflegekonzept wird aus gesundheitlicher Perspektive evaluiert, ob durch das Reko Verbesserungen in der Versorgung der Patienten erzielt werden können. Hierbei spielen insbesondere auch ethische Fragestellungen eine Rolle.

6 Ausblick

Ausgangspunkt des vorliegenden Beitrags war die Idee, defizitäre Krankenhäuser, die von einer Schließung bedroht sind, in IKT-basierte Pflegekompetenzzentren umzuwandeln. Zu diesem Zweck wurden zunächst eine Stakeholderanalyse *(FF1)* sowie eine systematische Anforderungsanalyse *(FF2)* durchgeführt. Aus der Literatur, verwandten Best Practice-Beispielen, Experteninterviews und Workshops wurden 36 Anforderungen abgeleitet. Auf dieser Basis wurde ein IKT-Instrumentarium erarbeitet, welches die Grundlage für das vorgestellte Reko-Konzept bildet *(FF3)*. Das Projektvorhaben im methodischen Rahmen eines Design Science Research-Ansatzes soll in der Modellregion Nordhorn durchgeführt und evaluiert werden. In der ersten Phase werden die Stakeholder sukzessive in den Räumlichkeiten des Marienkrankenhauses Nordhorn zusammengeführt und weitere individuelle Anforderungen werden erhoben. Anschließend wird das vorgestellte IKT-Konzept implementiert. Mit Hilfe der skizzierten Evaluation wird untersucht, inwiefern IKT dazu beitragen kann, die Pflegeversorgung in ländlichen, strukturschwachen Regionen zu verbessern. Abschließend werden Geschäftsmodelle erarbeitet, womit das Modellprojekt in andere Regionen transferiert werden kann. Limitierend ist anzumerken, dass die Anforderungen an ein Reko stets an regionale Bedürfnisse und vorhandene Strukturen angepasst werden müssen und somit nicht uneingeschränkt übertragbar sind. Im vorliegenden Beitrag wurde außerdem die Politik nicht als beteiligter Stakeholder berücksichtigt. Für eine Umsetzung in weiteren Regionen müsste diese jedoch gesetzliche Rahmenbedingungen schaffen und notwendige Finanzierungshilfen beisteuern.

In diesem Zusammenhang ergibt sich weiterer Forschungsbedarf aus ökonomischer, juristischer, ethischer Perspektive. Können Reko langfristig rentabel sein? Welche gesetzlichen Anpassungen müssen für eine erfolgreiche Umsetzung eines Rekos durchgeführt werden? Wie kann eine ethisch vertretbare Qualität der Pflege gewährleistet werden? Außerdem ist eine Untersuchung von Anreizsystemen notwendig, die zeigt, wie Patienten von der Nutzung einer Pflegeplattform überzeugt werden können. Sofern mit der Evaluation gemessen werden kann, dass die Pflegeversorgung sich durch die Imple-

mentierung des Rekos verbessert hat, kann das Pilotprojekt anderen defizitären Krankenhäusern als Vorbild dienen und auf weitere Regionen übertragen werden.

Literatur

[AHP17] Augurzky, B.; Hentschker, C.; Pilny, A.: Krankenhausreport 2017.

[An17] Anthony Berauk, V.L., Murugiah, M.K., Soh, Y.C., Chuan Sheng, Y., Wong, T.W., Ming, L.C.: Mobile Health Applications for Caring of Older People: Review and Comparison. Therapeutic Innovation & Regulatory Science 52, 374–382 (2017).

[At18] ATMoSPHÄRE, 2018. URL: https://www.atmosphaere.org. Abgerufen am 04.04.2019.

[BDO14] Deutsches-Krankenhaus-Institut; BDO: Ländliche Krankenhausversorgung Heute Und 2020. (2014).

[Be17] Beinke, J. H.; Meier, P.; Nickenig, H.-P.; Teuteberg, F.: Smart Home Predictive Analytics. In: INFORMATIK 2017 (2017).

[BHH18] Boll, S.; Hein, A.; Heuten, W.: Technologien für eine bedarfsgerechte Zukunft der Pflege. In: Zukunft der Pflege Tagungsband der 1. Clusterkonferenz 2018 S. 1.

[Br17] Breisig, T.; Felscher, A.; Hein, A.; Hülsken-Giesler, M.; Möller, W.; Erbschwendtner, S.; Fifelski, C.; Gilbert, J.; GLunz, L. M.; Isken, M.; Siemer, M.: Gesunde Pflege im Fokus - Entwicklung von demografiesensiblen, technikunterstützten Arbeitsprozessen in ambulanten Pflegeorganisationen - Das Projekt ITAGAP. Altmann, T., & Fuchs-Frohnhofen, P., Weimar, 2017.

[Ch09] Chenot, J.-F.: Cluster randomised trials: an important method in primary care research. In: Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 103 (Jan. 2009) 7, S. 475–480.

[DAG18] Deutsche-Alzheimer-Gesellschaft-e.V.: Mit Demenz im Krankenhaus, https://www.deutsche-alzheimer.de/angehoerige/mit-demenz-im-krankenhaus.html. Abgerufen am 01.04.2019.

[Di18] Dijkstra, N. E.; Sino, C. G. M.; Heerdink, E. R.; Schuurmans, M. J.: Development of eHOME, a Mobile Instrument for Reporting, Monitoring, and Consulting Drug-Related Problems in Home Care: Human-Centered Design Study. In: JMIR human factors 5 (2018) 1.

[FTI16] Frehe, V.; Teuteberg, F.; Ickerott, I.: IKT als Enabler für soziale Innovationen in Smart Rural Areas – Das Alter im ländlichen Raum hat Zukunft. In: Proceedings zur Multikonferenz Wirtschaftsinformatik (MKWI) 2016 (2016) May.

[Ge17] Gencer, D., Meffert, C., Herschbach, P., Hipp, M., Becker, G.: Belastungen im Berufsalltag von Palliativpflegekräften–eine Befragung in Kooperation mit dem Kompetenz Zentrum Palliative Care Baden-Württemberg (KOMPACT). Das Gesundheitswesen. (2017).

[Gu17] Guhra, M.: Demenz / Delir im Allgemeinkrankenhaus. (2017).

[Ha14] Haux, R.; Hein, A.; Kolb, G.; Künemund, H.; Eichelberg, M.: Five years of interdisciplinary research on ageing and technology: Outcomes of the Lower Saxony Research Network Design of Environments for Ageing (GAL) – An introduction to this Special

Issue on Ageing and Technology. In: Informatics for Health and Social Care (2014) 3–4, S. 161–165.

[HC10] Hevner, A.; Chatterjee, S.: Design research in information systems: theory and practice, vol. 22. Springer Science & Business Media, 2010.

[He04] Hevner, A.; March, S. T.; Park, J.; Ram, S.: Design science in information systems research. In: MIS Q 28 (2004), S. 75–105.

[HHF18] Hanke, F.-C.; Hochstadt, S.; Fröhndrich, N.: Kompetenznetz interdisziplinäre Geriatrie. In: MSD Gesundheit.

[IT18] ITAGAP. URL: http://itagap-projekt.de/. Abgerufen am 03.04.2019.

[JP16] Jelonek, M., Prilla, M.: Motivational Aspects of Using Augmented Reality Glasses in Care. In: Weyers, B., Dittmar, A. (Hrsg.) Mensch und Computer 2016 Workshopbeiträge 1–6. (2016).

[JP17] Jalaliniya, S., Pederson, T.: Qualitative Study of Surgeons Using a Wearable Personal Assistant in Surgeries and Ward Rounds. Ehealth 2016. LNICST 181, 208–219 (2017).

[KL15] Klein-hitpaß, U.; Leber, W.-D.; Scheller-kreinsen, D.: Strukturfonds: Marktaustrittshilfen für Krankenhäuser. 15 (2015), S. 15–23.

[KM18] Klie, T.; Monzer, M.: Regionale Pflegekompetenzzentren - Innovationsstrategien für die Langzeitpflege vor Ort. Beiträge zur Gesundheitsökonomie und Versorgungsforschung. 25th ed. Andreas Storm (DAK Gesundheit), Hamburg/ Freiburg, 2018.

[MN07] Myers, M. D.; Newman, M.: The qualitative interview in IS research: Examining the craft. In: Information and Organization 17 (2007) 1, S. 2–26.

[Mü12] Müller, C.; Neufeldt, C.; Randall, D.; Wulf, V.: ICT-development in residential care settings. In: Conference on Human Factors in Computing Systems Proceedings (May. 2012), S. 2639.

[Pe19] Peters, E.; Pritzkuleit, R.; Beske, F.; Katalinic, A.: Demografischer Wandel und Krankheitshäufigkeiten. In: Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz 53 (2010) 5, S. 417–426.

[Pr12] Prilla, M.; Frerichs, A.; Rascher, I.; Herrmann, T.: Partizipative Prozessgestaltung von AAL-Dienstleistungen: Erfahrungen aus dem Projekt service4home. In (Leimeister, J. M.; Shire, K. A. Hrsg.): Technologiegestützte Dienstleistungsinnovation in der Gesundheitswirtschaft. 2012.

[RK15] Reimers, K.; Klein, S.: Arzneimitteltherapiesicherheit im Spannungsfeld von vollständiger Medikationsübersicht, mündigem Patienten und individualisierter Medikation, 3rd ed. Cuvillier Verlag Göttingen, 2015.

[Se18] senimed-IT. URL: https://perspectiv.de/aerztenetze/senimed-it-erleichtert-die-pflege-und-hilft-bei-der-amts/. Abgerufen am 03.04.2019.

[Sp10] Spellerberg, A.: Intelligente Technik für das selbstständige Wohnen im Alter: Ambient Assisted Living für Komfort, Sicherheit und Gesundheit. Tagungsband der eHealth2010: Health Informatics meets eHealth. Schreier G, Hayn D, Ammenwerth E, Wien, 2010, S. Österreichische Computer Gesellschaft Nr. 264.

[WW02] Webster, J.; Watson, R. T.: Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: MIS Quarterly 26 (2002) 2, S. xiii–xxiii.

Extended Abstracts

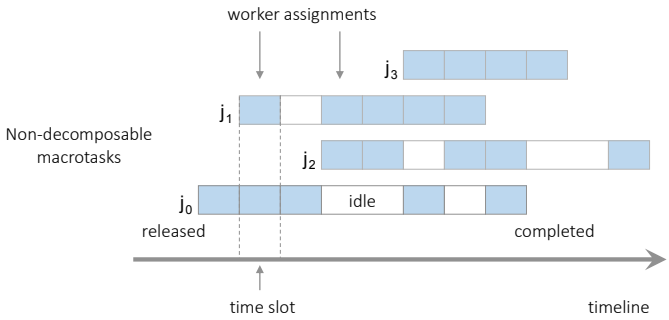
Online Sequencing of Non-Decomposable Macrotasks in Expert Crowdsourcing

Presentation of work originally published in ACM Transactions on Social Computing, Volume 1, Issue 1, Article No. 1, February 2018.

Heinz Schmitz[1], Ioanna Lykourentzou[2]

Keywords: crowdsourcing optimization; cooperative social computing; online scheduling decisions; macrotask scheduling

We introduce the problem of Task Assignment and Sequencing (TAS), which models online optimization in expert crowdsourcing settings that involve non-decomposable macrotasks. Non-decomposition is a property of certain types of complex problems, like the formulation of an R&D approach or the definition of a research methodology, which cannot be handled through the *divide and conquer* approach typically used in microtask crowdsourcing. In contrast to splitting the macrotask to multiple microtasks and allocating them to several workers in parallel, our model supports the sequential improvement of the macrotask one worker at a time, across distinct time slots of a given timeline, until a sufficient quality level is achieved. An online environment is assumed where expert workers are available only at specific time slots and worker/task arrivals are not known a priori.



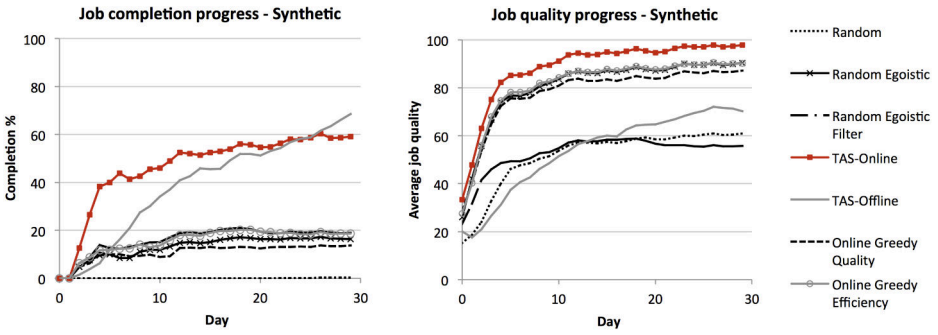
This new crowd work optimization model makes a conceptual shift from task decomposition and worker parallelization to task non-decomposition and worker sequencing. It combines

[1] Hochschule Trier, Schneidershof, 54293 Trier, Germany, h.schmitz@hochschule-trier.de
[2] Utrecht University, Princetonplein 5, 3584 CC Utrecht, The Netherlands, i.lykourentzou@uu.nl

doi:10.18420/inf2019_11

assignment decisions (per time slot and across tasks) with sequencing decisions (rolling out these assignments along a timeline) under reasonable constraints. We capture this setting with a mathematical formulation for TAS and prove its strong NP hardness.

With respect to this setting, we propose TAS-ONLINE, an online algorithm that aims to complete as many tasks as possible within budget, required quality and a given timeline, without any future input information regarding job release dates or worker availabilities. We illustrate, through synthetic and real-world experiments, that TAS-ONLINE can achieve more completed jobs, lower flow times and higher quality compared to five typical benchmarks.



Our small-scale real-world experiment on CrowdFlower[3] was intended to provide an initial and qualitative viewpoint of the model's performance in a real-world setting. The task we used was collaborative news article writing, where workers from an initial hiring pool were asked to build on each other's content sequentially, enriching a news article text on a given topic. At the end of the scheduling period, the competing benchmark algorithm achieved a successful completion of 3 out of 6 jobs, while our optimization algorithm successful completed 5 jobs. As it was expected the benchmark algorithm either allowed workers of the minimum necessary expertise to take a job, thus delaying the job's quality progress too much, or it starved the job of budget. On the other hand TAS-ONLINE selected workers in such a way as to improve job completion within the given time period.

Our results have implications for enhancing the Quality of Service of crowdsourcing platforms offering non-decomposable complex tasks, but also for allowing online expert crowdsourcing communities to make better use of their human capital and available expertise. Multiple future extensions can be foreseen. These include extending the proposed TAS model to handle requirements such as budget flexibility, the non-acceptance of assignments by the workers, different job quality aggregation mechanism, learning, varying modes of assignment order and number of assignments per worker, as well as forecasting.

[3] http://www.crowdflower.com

The Practice Turn in IT security – An Interdisciplinary Approach

Laura Kocksch[1] Andreas Poller[2]

Abstract: IT security has traditionally been approached as an isolated technological phenomenon or as a matter of user incompetency. In our work, we suggest to apply a practice turn to the study of IT security to unfold the nexus of practices that is involved in engaging with IT security work. In doing so, IT security becomes an organizational, social and political phenomena that demands interdisciplinary attention from computer science and social science alike.

Keywords: IT security; social science; practice; organization

IT security has traditionally been perceived as a matter of technological precision and function. Recent events have demonstrated the effects of insecure IT systems for organizations, businesses and society as a whole. Hacks are socially and politically motivated and solely technological explanations do not suffice.

Because of its vital role, IT security is studied in computer science and engineering but also in interdisciplinary collaborations. For instance computer scientists and psychologists test technological security measures for their compliance with users' needs and competencies. While the usability of security measures has been in the focus of attention, recently, scholars have argued developers of software products need to be educated and need to comply too. By the same token, IT security is scripted into legal frameworks, standards are being set up and education resources have developed. We build on this impulse to unpack IT security as more than a technological failure or a question of users and software developers' in-/competency, and argue that we need to gain a broader understanding of what assembles IT security in social practice. To this end, we strive to account for the complex webs of practices that technologies are involved in and that insecurities can have potential effects on.

When turning our attention to the daily practices involved in engaging in security work, we shed light on its ambivalence and messiness. By taking the practice turn, we shift perspectives away from individual actors or social groups to situated actions and the heterogeneous stakeholders involved. We emphasize on the tense negotiation, collaboration and cooperation involved in situated security work, and ask how IT security is constantly enacted, maintained, contested and cared for in dispersed practices.

[1] Ruhr University Bochum, Faculty of Social Science, Universitätsstrasse 150, 44801 Bochum, Germany laura.kocksch@rub.de

[2] Fraunhofer Institute for Secure Information Technology, Rheinstrasse 75, 64285 Darmstadt, Germany, andreas.poller@sit.fraunhofer.de

doi:10.18420/inf2019_12

Practice-theoretical approaches in social science shift the attention to the assemblages of heterogeneous actors that contributed to a given situation. Turning to practice emphasizes on the organizational, social and material constraints of daily life. This turn has not yet been applied to the study of IT security. In doing so, we suggest to describe IT security as distributed across heterogeneous actors; not just individual developers or users.

We exemplify the practice turn in IT security in two empirical cases that we draw from a one-year-study at a large German software vendor in 2016. The first case stresses the *material practice of IT security in organizations* [Po17]. We investigated events in the aftermath of a software penetration test whose results the software developers translated into single defect reports managed by means of an issue tracking system. We realized that the issues themselves had agency. The issues demanded a specific way of taking care of security that complied with the company's business strategy. Had security taken on a different material than issues it could have been introduced more fundamentally in the software product. But as an *issue* it was merely one task to check of a list.

The second case emphasizes the *morality and politics of IT security* [Ko18]. We observed a security training using team-ethnographic methods with a particular focus on the interaction between a security consultant and a team of developers. We soon realized that the team comprised highly trained and expert developers. However, during the training, *bad* programming skills were blamed for the findings of the penetration test. The training session was conceptualized as a one time event that should scare and blame developers, while letting unnoticed the myriad of material, management and collaborative practices that contribute to a software product. IT security was defined as a shortcoming in developers' actions; clearing all other actors in the company of responsibility. The training located security at developer's hand framing political responsibility as an individual software development problem.

By applying the practice lens to the study of IT security in our two examples, we demonstrate that IT security demands *careful* intervention that is sensitive to the situated practices of developers. IT security is an issue for society, business and politics alike, and because of its significance, we suggest that it should not be reduced to isolated actors but acknowledged as a dispersed phenomena, entangled with heterogeneous practices. Further interdisciplinary research between computer scientists and social scientists is needed in this area.

Literaturverzeichnis

[Ko18] Kocksch, Laura; Korn, Matthias; Poller, Andreas; Wagenknecht, Susann: Caring for IT Security: Accountabilities, Moralities, and Oscillations in IT Security Practices. Proc. ACM Hum.-Comput. Interact., 2(CSCW):92:1–92:20, November 2018.

[Po17] Poller, Andreas; Kocksch, Laura; Türpe, Sven; Epp, Felix Anand; Kinder-Kurlanda, Katharina: Can Security Become a Routine?: A Study of Organizational Change in an Agile Software Development Group. In: Proc. CSCW '17. ACM, New York, NY, USA, S. 2489–2503, 2017.

Program Comprehension and Developers' Memory

Presentation of work originally published in the Proc. of the 40th Intl. Conf. on Software Engineering

Jacob Krüger[1,2] Jens Wiemann[2] Wolfram Fenske[2] Gunter Saake[2] Thomas Leich[1,3]

Abstract: In this extended abstract, we summarize our paper "Do You Remember This Source Code?", published at the International Conference on Software Engineering 2018 [Kr18]. We discuss implications of our results on forgetting in the context of program comprehension, providing a more contextual perspective on our results compared to the original paper and a previous abstract [Kr19].

Keywords: Familiarity; forgetting; empirical study; maintenance; program comprehension

Software developers constantly design, implement, maintain, and re-engineer software systems. For this purpose, they have to understand the program itself, which is the most time-consuming and costly activity that software developers perform—called *program comprehension*. There exist numerous studies to show the impact of various factors on program comprehension, and techniques to facilitate this activity, for instance, based on comments or identifier names. However, few studies investigate the remaining *software familiarity* of developers with a system. This familiarity comprises knowledge about the code, design, architecture, and application of the software that is *forgotten* over time. While some techniques (e.g., expertise identification) consider factors related to forgetting (i.e., time, authorship), we are not aware of empirical analyses that show the impact of these factors on developers' memory.

We conducted an empirical study on the self-assessed familiarity of 60 open-source developers with source code they worked on [Kr18]. More precisely, we investigated whether the well-known forgetting model of Ebbinghaus applies to software familiarity, analyzed three factors (i.e., number of edits, ratio of own code, tracking behavior) from learning theory that should also influence forgetting, and computed the memory strength of our participants. In Table 1, we show the results of testing the significance of our observations. To this end, we used two rank correlations (Kendall's Tau is more strict than Spearman's Rho) that indicate the effect each factor has on our participants' memory. We remark that we controlled for the file sizes to ensure that these had no impact, as we asked our participants to state their remaining familiarity as ratios of the files.

[1] Harz University of Applied Sciences Wernigerode; tleich@hs-harz.de
[2] Otto-von-Guericke-University Magdeburg; jkrueger@ovgu.de; wfenske@ovgu.de; saake@ovgu.de
[3] METOP GmbH Magdeburg

doi:10.18420/inf2019_13

As we can see in Table 1, the file sizes did not correlate with familiarity, which was the prerequisite for our analysis. The remaining results showed a local maximum in familiarity after approximately 120 days had passed since the last edit of a file. This highlights the impact of factors other than time on familiarity. Our results indicate that the *number of edits*, which represents repeated commits to the same file at different days, has a moderate to strong, positive correlation to familiarity. For the *ratio of own code* that a developer implemented, we also found a moderate, positive correlation. Consequently, we assume that repeatedly working on a file and implementing more of its code does indeed improve a developer's memory of that file. In contrast, we found no correlation for the *tracking behavior*, which refers to developers following and analyzing the changes others employ on the file. However, this may be due to different understandings of tracking and we need to conduct further analyses to see whether this finding holds true. Finally, we found that the averaged curve resembles that of Ebbinghaus, especially if we remove the number of edits as factor. Thus, Ebbinghaus' forgetting model seems to apply to software engineering, but only if developers implemented a file in one session. Overall, we argue that we need an adapted forgetting curve for software development that considers additional factors that are specific to the activities, processes, and interactions of developers.

Tab. 1: Spearman's Rho (r_s), Kendall's Tau (τ), and the corresponding significance (sig.) values.

Factor	r_s	sig.	τ	sig.
File Size	0.16	0.22	0.11	0.24
Number of Edits	0.67	4.56×10^{-9}	0.55	5.18×10^{-8}
Ratio of Own Code	0.55	4.57×10^{-6}	0.42	6.86×10^{-6}
Tracking Behavior	0.04	0.79	0.02	0.8

We conducted this research in the context of re-engineering, for which developers first need to comprehend the software. As program comprehension is costly, the costs for re-engineering heavily depend on the remaining familiarity a developer has. In practice, we can find various scenarios in which the question arises who should perform a task. For example, consider a developer who implemented a file some time ago. Other developers made changes to this file, introducing new functionality or fixing bugs. The question is, who is best suited to perform a new task on this file? It may be the original developer (ratio of own code), the one who did most changes (number of edits) or the one with the most recent changes (time). This question is impossible to answer precisely, but our research indicates that existing heuristics will benefit from taking the impact of forgetting on software familiarity and program comprehension into account.

Acknowledgments: Supported by DFG grants LE 3382/2-1 and SA 465/49-1.

References

[Kr18] Krüger, Jacob; Wiemann, Jens; Fenske, Wolfram; Saake, Gunter; Leich, Thomas: Do You Remember This Source Code? In: ICSE. 2018.

[Kr19] Krüger, Jacob; Wiemann, Jens; Fenske, Wolfram; Saake, Gunter; Leich, Thomas: Understanding How Programmers Forget. In: SE/SWM. 2019.

The Use of Design Thinking for Requirements Engineering: An Ongoing Case Study in the Field of Innovative Software-Intensive Systems

Presentation of work originally published in the Proceedings of the 26th IEEE International Requirements Engineering Conference, Banff, Canada 2018

Jennifer Hehn[1], Falk Uebernickel[2]

Keywords: Design Thinking, Requirements Engineering, Innovative Software-Intensive Systems

1 Extended Abstract

Effective Requirements Engineering (RE) is recognized to be one of the most crucial activities in software-intensive development projects. However, practitioners and scholars have also revealed its numerous challenges. Especially nowadays, when software development increasingly calls for agile and human-centered practices to address the often fuzzy needs of the various stakeholders involved. The popular approach of Design Thinking (DT) has gained recognition as a way to approach product and software development with an interdisciplinary team, qualitative user research methods, rapid (non-technical) prototyping techniques, and iterative learning cycles. This diverging way of problem-solving is notably different from the rather converging and more formal RE practices. We postulate that the strongly human-oriented working mode of DT can complement the more formal, technology-driven RE practices. This is a new field of exploration that has not been systematically examined yet. To enhance this understanding with empirical evidence we set up a longitudinal case study in an agile development setting from idea conceptualization to market-ready implementation. We investigated a software-intensive development project over a time frame of 1.5 years in a large utility company in Europe. The objective of the project was to design a digital platform in the energy sector. The project management decided to apply DT and (later) Scrum to better understand the problem domain before drawing conclusions too early. The interdisciplinary project team included domain experts, user researchers, business specialists, IT and technology experts, and a project lead. Table 1 shows the main project phases of (1) exploration, (2) alpha prototyping, and (3) market launch and their respective objectives, activities, roles, and outcomes.

[1] University of St.Gallen, Institute of Information Management, Müller-Friedberg-Strasse 8, 9000 St.Gallen, jennifer.hehn@unisg.ch
[2] University of St.Gallen, Institute of Information Management, Müller-Friedberg-Strasse 8, 9000 St.Gallen, falk.uebernickel@unisg.ch

doi:10.18420/inf2019_14

Phase	***Exploration***	***Alpha Prototyping***	***Market Launch***
Duration	3 months, 8 FTE	7 months, 15 FTE	4 months, 22 FTE
Objective	Understand the problem and create a product vision	Develop a functional alpha prototype	Market launch of the platform
Activities	DT as guiding process: (1) empathize, (2) define, (3) ideate, (4) prototype, and (5) test in several iterations	Scrum as guiding framework; DT tools to enhance communication and ideation with stakeholders	Scrum as guiding framework; enhanced priority on de-fining the go-to-market strategy; friendly user tests
Roles	Each team member is involved into all activities to elicit needs and requirements	Onboarding of development team; business model focus; DT team transitions into Product Owner role	Software development team is extended; split between technical, business model, and Product Owner role
Outcome	Mockup with core functionalities; customer journeys define context of use	Proof of Concept demonstrates feasibility and viability of the solution	Minimum Viable Product (MVP) is ready for market entry
Conclusion	DT as guiding process; requirements elicitation is a sequence of team-based efforts	DT used as a toolbox; Scrum is guiding framework; part of DT team becomes Product Owner	DT as a mindset; Selected DT tools provide support, yet agile development practices dominate

Tab. 1: Project Overview (adapted from [HU18])

We found that DT has the ability to address some of the known challenges in agile RE, e.g., to expose tacit knowledge of stakeholders through prototypes. In particular, we learned these lessons by using DT to elicit requirements and define a platform solution:

- *DT provides a structured process to requirements elicitation for wicked problems.* DT offers a prescriptive guideline to apply methods which are commonly used in RE to elicit stakeholder needs and requirements.
- *DT leverages a team-based effort for requirements elicitation.* The role of the Product Owner is inhabited by an interdisciplinary DT team leading to a comprehensive requirements elicitation through various viewpoints.
- *DT emphasizes the elicitation of user requirements with a special focus on usability.* DT puts priority on discovering requirements related to usability, workflows tasks, and user interface.
- *DT supports a seamless integration of upfront and concurrent RE practices.* DT evolves from an upfront definition of stakeholder needs, into tool support with human-centered principles that, both, link well to common agile practices.

DT offers several possibilities for coping with today's (agile) RE challenges. We see potential to combine DT and RE in many dimensions to help create human-centered software solutions more effectively. We consider DT as an "extended arm" for RE to approach wicked problems and explore actual needs upfront with a prescriptive process, while RE provides an integration framework for DT into later staged software development life cycles. However, in our study we also found that DT must also account for a number of anomalies that are not clarified yet, similar to agile practices. We plan to analyze challenges like the high dependency of people, traceability difficulties, or the lack of formalization. Here, we plan to learn from the more mature discipline of RE.

Towards Ubiquitous Requirements Engineering

Presentation of work originally published in the Proc. of the 26th IEEE International Requirements Engineering Conference

Karina Villela, Anne Hess, Matthias Koch, Rodrigo Falcão, Eduard Groen, Joerg Doerr[1]
Carol Valero and Achim Ebert[2]

Abstract: We have perceived barriers that prevent requirements engineers from contributing properly to the development of the software systems that underpin the digital transformation. We have also realized that breaking down each of these barriers would contribute to requirements engineering (RE) becoming ubiquitous in certain dimensions. In this paper, we point out the transformation that is required to break down each barrier and briefly discuss each dimension of ubiquity. Our goal is to raise the interest of the research community in providing approaches to address the barriers and move towards ubiquitous RE.

Keywords: distributed RE; RE with the society; RE for IoPTS; CrowdRE; RE for ecosystems

1 Dimensions of Ubiquity and Required Transformations

RE Everywhere (from geographic colocation to worldwide distribution): While companies can decide about the geographic location of their units and about whether or not to outsource their software development, partners in a strategic digital ecosystem may be located anywhere in the world. Requirements engineers should be capable of conducting RE activities with end users remotely and even asynchronously, depending on the time zone. We envision the development of software environments to support Virtual RE based on augmented reality and motivation mechanisms.

RE with Everyone (from wishing for experienced end users to empowering newbies): In smart rural areas, for example, the end users are mainly citizens who have never contributed to the development of a software system and have no special technology affinity. We envision both the adaptation of Participatory Design techniques to empower citizens to actively participate in the process of digital transformation of their cities and the development of a framework to support requirements engineers in finding the appropriate fit between end users' characteristics and RE methods.

[1] Fraunhofer IESE, Fraunhofer-Platz, 67663 Kaiserslautern, Germany <vorname.nachname>@iese.fraunhofer.de
[2] TU Kaiserslautern, Computer Graphics & HCI, 67663 Kaiserslautern, Germany <nachname>@cs.uni-kl.de

doi:10.18420/inf2019_15

RE for Everything (from focusing on software to holistically taking into consideration people, things, and services): Nowadays, people, things, and web services can have a digital identity and be interconnected via the Internet (IoPTS). This requires a completely different way of understanding the context and the components of a software solution, due to the nature of the entities. A Digitalization Potential Analysis could be a top-down, human-based approach for defining the vision of the software solution, whereas automatic context modeling followed by the derivation of context-aware requirements would be a bottom-up, automated approach.

Global Automation (from supporting direct interaction with a set of representative end users to also allowing a crowd to indirectly provide requirements): In some situations, the group of users might be so heterogeneous that it might be easier and more effective to allow anyone to provide potential requirements rather than to try to identify a set of representative end users or think about appropriate personas. Crowd RE stands for performing RE with the support of a crowd of stakeholders in an automated way through two complementary mechanisms: User Feedback Analysis and Usage Mining. Both mechanisms aim at extracting requirements for the evolution of the target software and allow stakeholders to contribute requirements without being aware of it.

Openness (from wishing for well-understood processes and groups of end users to accepting openness) & Cross-Domain (from dealing with one domain at a time to dealing with multiple domains): The business value of a software ecosystem relies on new processes made possible only by orchestrated cooperation among partners. However, concrete partners might still be unknown or might change; even if the actual partners are known, their contributions to the ecosystem might still be unclear/undecided. In this context, groups of end users and processes tend to be not well understood and requirements – in the sense of a perceived need for a functionality or quality attribute – might not exist. Furthermore, a software ecosystem is, by nature, cross-domain, as partners from different sectors with different services decide to combine their strengths and offer upper-level services. Approaches for performing RE for software ecosystems are under development. Creativity techniques are expected to play an important role, while software tools can support the search for the right partners and the mapping of relevant concepts across domains. In any case, the skills of requirements engineers need to shift from being able to elicit requirements to being able to propose requirements and thereby strongly support stakeholders in shaping the ecosystem vision.

These dimensions of ubiquity provide an overview of the future directions for RE from our mixed practitioner / researcher perspective. For more details, see the full paper [Vi18], where references to related work in each dimension are provided.

References

[Vi18] Villela, K; Hess, A; Koch, M; Falcão, R; Groen, E; Doerr, J; Valero, C; Ebert, A: Towards Ubiquitous RE: A Perspective on Requirements Engineering in the Era of Digital Transformation. In: International Requirements Engineering Conference. IEEE, pp. 205–216, 2018.

Track 2 – Internet of Everything

Internet of Everything

Anna Förster[1], Matthias Wählisch[2]

The Internet is more than a communication infrastructure for a selected group of people. It is an integral part of our daily life. In this track, we focus on all research questions that relate to the interconnection of users and devices as well as machine-to-machine communication. This includes common topics such as network protocols and algorithms for the Internet of Things but also emerging fields such as the intercon-nection of cars and factories.

In detail, the topics of interest for this track included: Internet of Things; industry 4.0; car2X communication; M2M communication; security and privacy; 5G, 6TSCH, CoAP, MQTT, etc.; applications; services; protocols; network architectures; deployment experiences and open challenges; standardization.

The program committee consisted of

- Martina Brachmann, RISE
- Torsten Braun, Universität Bern
- Falko Dressler, Universität Paderborn
- Anna Förster, Universität Bremen (co-chair)
- Elena Gaura, Coventry University
- Ulf Kulau, TU Braunschweig
- Olaf Landsiedel, Universität Kiel
- Ramona Marfievici, Nimbus Research Centre,Cork Institute of Technology
- Kay Roemer, TU Graz
- Jochen Schiller, FU Berlin
- Thomas Schmidt, HAW Hamburg
- Ralf Steinmetz, TU Darmstadt
- Matthias Wählisch, Freie Universität Berlin (co-chair)
- Lars Wolf, TU Braunschweig

[1] Universität Bremen, anna.foerster@comnets.uni-bremen.de
[2] Freie Universität Berlin, m.waehlisch@fu-berlin.de

doi:10.18420/inf2019_16

The track „Internet of Everything“ received eleven submissions. After a single-blind peer review process, we selected six papers for presentation. One paper is an extended abstract, which summarizes previously published work at the IEEE INFOCOM Workshop on Ultra-Low Latency in Wireless Networks 2019.

The selected papers span the variety of the Internet of Everything. They report about a cross-layer pacing approach; scalable secure IoT network integration; the potentials of the secure personal health record; context maps for connected cars; a structured comparison of blockchain and distributed ledger technologies; and the combination of semantic data integration and edge computing.

The track chairs gratefully acknowledge the work of all authors and the time and dedication of the technical program committee to shape the program. We explicitly appreciate the support of the GI organizing committee, in particular we would like to thank Kurt Geihs and Martin Lange for their continuous support while organizing the track and compiling the proceedings.

Full Papers

Die elektronische Gesundheitsakte als Vernetzungsinstrument im Internet of Health

Anwendungsfälle und Anbieter im deutschen Gesundheitswesen

Christian Fitte[1], Pascal Meier[1], Alina Behne[1], Dafina Miftari[1] und Frank Teuteberg[1]

Abstract: Das Internet of Everything bietet große Potenziale, die Gesundheitsversorgung zu verbessern und die Grundlage für ein vernetztes Internet of Health (IoH) zu bilden. Während in den letzten Jahren viele digitale Insellösungen entstanden sind, mangelt es im Gesundheitswesen an einer intelligenten Verknüpfung von Personen, Prozessen, Daten und Dingen. Im vorliegenden Beitrag wird elektronische Gesundheitsakte (eGA) als patientenzentriertes Vernetzungsinstrument im IoH vorgestellt. Für eine Analyse des State of the Art werden zunächst aktuelle Anbieter einer eGA in Deutschland vorgestellt und 25 Anwendungsfälle der eGA identifiziert. Anschließend wird das Potenzial der eGA als Vernetzungsinstrument im IoH herausgearbeitet. Im Rahmen von neun Experteninterviews mit Gesundheitsdienstleistern werden Anwendungsfälle der eGA sowie Herausforderungen für den flächendeckenden Einsatz der eGA abgeleitet.

Keywords: Elektronische Gesundheitsakte, Vernetzung, Gesundheitswesen, Internet of Health.

1 Einleitung

Das Internet of Things (IoT) wird als eine vielversprechende Zukunftslösung beschrieben. Besonders im Gesundheitswesen haben IoT-basierte Technologien das Potenzial, die Versorgung zu verbessern und für Patienten zu vereinfachen [BEE18], [Ro18]. Neben der Verknüpfung von Dingen ist im Gesundheitsbereich die Vernetzung der beteiligten Akteure, der Prozesse sowie der dazugehörigen Daten von Bedeutung. Diese vier Bestandteile werden in dem Internet of Everything (IoE) zusammengefasst [Mi15]. Die Patientendaten können über verschiedene Geräte und Sensoren erfasst, über Anwendungen verarbeitet und über ein Benutzerendgerät ausgegeben werden wie bspw. die Überwachung von Vitalparametern mittels biomedizinischer Sensoren und dem anschließenden Teilen mit einem Facharzt. Diese Verknüpfung von Daten, Prozessen, Personen und Dingen kann im Gesundheitswesen durch die Anbindung an das Internet als Internet of Health (Things) und als Ausprägung des IoE verstanden werden [Ro18].

Aktuell besteht bereits eine Vielzahl an Insellösungen, welche verschiedene Gesundheitsdaten über das Internet kommunizieren. Bislang fehlt jedoch eine zentrale Plattform, die alle Bestandteile des Internet of Health (IoH) miteinander vernetzt. Die elektronische

[1] Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49076 Osnabrück, {christian.fitte; pascal.meier; alina.behne; dmiftari; frank.teuteberg}@uni-osnabrueck.de

doi:10.18420/inf2019_17

Gesundheitsakte (eGA) könnte diese zentrale Lösung darstellen und so als patientenorientiertes Vernetzungsinstrument fungieren. Sie verknüpft alle Gesundheitsakteure miteinander und erfüllt die vielseitigen Anforderungen, wie Ressourceneffizienz, Daten- und Prozessintegration und die sektorenübergreifende Vernetzung von Akteuren des Gesundheitswesens [BFT19]. Es werden zahlreiche Vorteile erzielt: Eine zentrale Speicherung von Diagnosen, Medikamenten- und Therapiepläne stärken die intersektorale Zusammenarbeit im Gesundheitswesen und beugen Doppeluntersuchungen sowie unerwünschten Medikamenteninteraktionen vor [Am16]. Arbeitsprozesse können effizienter gestaltet werden, da die administrative Belastung sinkt und Ärzte auf professionelle Entscheidungsunterstützungssysteme zurückgreifen können. In der Folge steigt die Qualität der Gesundheitsversorgung, da Ärzte mehr Zeit für die Patientenversorgung haben und Patienten durch eine aktive Einbindung besser motiviert werden können, ihre Therapiemaßnahmen einzuhalten. Zudem können aggregierte Gesundheitsdaten in anonymisierter Form wertvolle Erkenntnisse für die medizinische Forschung bringen [Mc15]. Die hohe Bedeutung der eGA spiegelt sich in dem geplanten Terminservice- und Versorgungsgesetz (TSVG) wider: Krankenkassen sind bis 2021 verpflichtet, gesetzlich Versicherten elektronische Patientenakten nach den Interoperabilitätsvorgaben der gematik zur Verfügung zu stellen [Kr18]. Daher etablieren sich zunehmend private Anbieter einer elektronischen Gesundheitsakte auf dem deutschen Markt. Für 60% der Deutschen sind die Vorteile überzeugend, sodass sie eine eGA nutzen würden, jedoch setzt sich aktuell kein Anbieter für den flächendeckenden Einsatz durch [RJ17]. Eine Aufarbeitung der Nutzungsmöglichkeiten der eGA als Vernetzungsinstrument im IoH fehlt nach dem heutigen Stand. Daher lassen sich folgende Forschungsfragen ableiten:

FF 1: Welche Anwendungsfälle können aktuell durch die eGA in Deutschland unterstützt und ermöglicht werden?

FF 2: Welche Möglichkeiten bietet die eGA als Vernetzungsinstrument im IoH?

FF 3: Welche Herausforderungen bestehen für den flächendeckenden Einsatz von eGAs und wie können diese überwunden werden?

Zur Beantwortung dieser Forschungsfragen wird ein multimethodisches Vorgehen angewendet. Zunächst wird die Entwicklung der eGA sowie die Abgrenzung zur elektronischen Patientenakte beschrieben. Mit Hilfe einer Marktrecherche wird ein Überblick über die derzeitigen eGA-Anbieter in Deutschland gegeben. Durch eine strukturierte Literaturrecherche werden Anwendungsfälle der eGA identifiziert (FF1). In Kapitel drei werden die vier Bestandteile des IoE im Gesundheitswesen vorgestellt. Anschließend wird die eGA als zentrales Vernetzungsinstrument im IoH vorgeschlagen (FF2). Anhand von neun Experteninterviews wird das Potenzial der eGA als Vernetzungsinstrument evaluiert sowie Herausforderungen für den flächendeckenden Einsatz identifiziert (FF3). Nach einer Diskussion der Ergebnisse folgen ein zusammenfassendes Fazit sowie ein Ausblick auf die zukünftige Entwicklung.

2 Die elektronische Gesundheitsakte im deutschen Gesundheitswesen

2.1 Entwicklung und Abgrenzung

In der Literatur und Praxis bestehen zahlreiche unterschiedliche Formen und Bezeichnungen von elektronischen Gesundheitsakten [HSN08], [Pr01]. Insbesondere im Vergleich internationaler Literatur fällt auf, dass z.T. verschiedene Bezeichnungen für gleiche Konzepte, aber auch gleiche Begriffe für unterschiedliche Konzepte verwendet werden [Ha17]. Daher legt dieses Kapitel dar, was im Rahmen dieses Beitrags unter dem Begriff eGA verstanden wird. Zunächst sind Patientenakten und Gesundheitsakten voneinander abzugrenzen. Patientenakten werden von Leistungserbringern geführt und implizieren, dass eine Krankheit vorliegt, bzw. eine Behandlung in einer Institution des Gesundheitswesens stattfindet [Am16]. Unter der ursprünglichen Form einer elektronischen Patientenakte (ePA) wurde die digitale Speicherung (z. B. durch Einscannen) von dokumentationspflichtigen Befunden innerhalb einer Institution verstanden (interne elektronische Patientenakte). Mit zunehmender Vernetzung sollten schließlich auch einrichtungsübergreifend Patientendaten in einer elektronischen Akte abgelegt werden (einrichtungsübergreifende elektronische Patientenakte) [HB17]. Im angelsächsischen Sprachraum wird die ePA als „electronic health record“ (EHR) bezeichnet [Am05].

Gesundheitsakten hingegen werden nutzerseitig vom Patienten selbst verwaltet und sind umfangreicher. Sie können jegliche Gesundheitsdaten, wie z. B. einrichtungsübergreifende Diagnosen, Medikations- und Therapiepläne, Impfpässe und weitere Dokumente enthalten. Im Gegensatz zu einer Patientenakte, die voraussetzt, dass der Nutzer tatsächlich erkrankt ist, können in Gesundheitsakten auch nicht professionell medizinische Daten, sog. „Wellnessdaten“, eingebunden werden, welche auch der Patient selbst hinterlegen kann, z. B. kontinuierlich gemessene Vitalparameter [Pr01]. Der Nutzer kann fallweise festlegen, welchem Akteur (Ärzten, Krankenkassen, Apotheken, u. a.) er Zugriff auf ausgewählte Daten gewährt. Der Zugriff kann auch im Rahmen einer telemedizinischen Behandlung erfolgen. Die Inhalte der Akte können in anonymisierter Form von Forschungszentren verwendet werden [Mc15]. International wird die eGA „personal health record“ (PHR) genannt. Gegenstand dieser Untersuchung ist die eGA.

2.2 Anbieter von elektronischen Gesundheitsakten in Deutschland

Um einen Überblick über aktuelle Lösungen von eGAs auf dem Markt zu geben, wurde eine Onlinerecherche durchgeführt. Die Auswahlkriterien hierfür bietet einerseits die oben genannte Definition einer elektronischen Gesundheitsakte und andererseits die Verbreitung sowie das Nutzerpotenzial. Es wurden sieben Anbieter einer eGA im Einsatz oder in einer produktiven Testversion identifiziert, die nachfolgend vorgestellt werden: ***Vitabook*** ist ein Online-Gesundheitskonto in Kooperation mit der Microsoft Cloud Deutschland, losgelöst von den Krankenkassen. Mit der Versicherungsnummer erhält

der Nutzer Zugang zu seinem persönlichen Konto. Von der Arztsuche und Terminbuchung bis zur Dokumentation von Vitalparametern bietet Vitabook ein großes Portfolio an möglichen Anwendungsfällen [Vi19]. Zudem bedient Vitabook den Pflegebereich mit der Software *ordermed*. Diese verbindet Pflegeheime und Pflegedienste mit Apotheken und Ärzten, um die Rezept- und Medikamentenbesorgung zu vereinfachen. Eine weitere elektronische Gesundheitsakte ist ***Vivy***, die am 17. September 2018 gefördert durch 21 gesetzlichen und vier privaten Krankenversicherungen startete. Mithilfe von Vivy soll vor allem die Verbindung zwischen Nutzern und Ärzten, Krankenhäusern, Laboren, Krankenkassen sowie Versicherungen gestärkt werden. Angeführte Funktionen sind automatische Wechselwirkungstests beim Scannen des Medikamentenpackungscodes, Gesundheitschecks und die Kopplung mit Fitnesstrackern. Aufgrund einer Initiative der Technikerkrankenkasse (TK) und IBM wurde die Gesundheitsakte ***TK Safe*** entwickelt, die sich seit April 2018 im Testbetrieb mit Betatestern befindet [TK19]. Insgesamt werden über die TK ca. 10 Millionen Versicherte erreicht. Die AOK gestaltetet in Kooperation mit Vivantes und Sana ein ***digitales Gesundheitsnetzwerk*** mit eGA für potenziell 26 Millionen Versicherte. Darin wird ebenfalls eine patientenzentrierte, nachhaltige Datenübertragung über die einzelnen Gesundheitsakteure angestrebt. Die niederländische eGA ***Forecare*** besteht seit 2006 und gehört seit 2017 zu Philips. Sie verspricht verschiedene Patientendaten nahtlos zwischen den bestehenden Systemen auszutauschen und wirbt mit der Einhaltung offener Standards sowie mit Zertifizierungen [Fo19]. ***HealthVault*** ist eine eGA des US Technologiekonzerns Microsoft, die eine Verknüpfung zu Applikationen und Sensoren ermöglicht, sodass aufgenommene Daten direkt zu HealthVault übertragen werden. Des Weiteren ist die Freigabe der persönlichen Daten an beliebige Personen möglich [He19]. ***PatientAssist*** gehört zu dem Unternehmen Healthcare X.0. Dabei soll dem Patienten, bspw. durch die erleichterte Erstellung von Tagebucheinträgen und die Verknüpfung zu Fitnesstrackern, die Datenlast abgenommen werden. Dieser Anbieter hebt sich mit PatientAssist durch die Möglichkeiten ab, bei Bedarf ein Notfallsignal an eine ausgewählte Vertrauensperson zu senden sowie bei Interesse und Einwilligung des Nutzers die eigenen Gesundheitsdaten für Forschungszwecke freizugeben [Pa19]. Im weiteren Verlauf des Beitrags werden die sieben vorgestellten Anbieter detailliert auf verschiedene Anwendungsfälle einer eGA untersucht.

2.3 Anwendungsfälle der elektronischen Gesundheitsakte

Zur Beantwortung der FF1 wurde eine systematische Literaturrecherche durchgeführt. Die Datenbanken EBSCOhost, Emerald, IEEEXplore, Medline, ProQuest, ScienceDirect, Scopus, Wiley and Google Scholar wurden mit dem Term (*EHR OR EPA OR EGA OR PHR) AND ("Use Cases" OR Anwendungsfälle)* durchsucht. Damit wurden die 25 Anwendungsfälle identifiziert, welche mit ihrer jeweiligen Quelle in Tab. 1 aufgeführt sind. Gleichzeitig wird ein Überblick gegeben, welcher der oben genannten Anbieter laut Anwendungsbeschreibungen auf der Produktwebseite die jeweiligen Anwendungsfälle unterstützt. Aus diesen Anwendungsbeschreibungen wurden weitere Anwendungsfälle identifiziert, die in der Tab. 1 ergänzt wurden.

Tab. 1: Anwendungsfälle der elektronischen Gesundheitsakte

Literatur									Anwendungsfälle		Anbieter							
Ambinder (2005)	Amelung et al. (2016)	Haas (2017)	Heinze et al. (2017)	Hogan et al. (2011)	McCowan et al. (2015)	Monica (2017)	Neuhaus et al. (2006)	Schwarze et al. (2005)			Digitales Gesundheitsnetzwerk	Forcare	Microsoft Health Vault	Patient Assist	TK-Safe (TK & IBM)	Vitabook	Vivy	WebMD Personal Health Manager
Kommunikation																		
x	x	x	x				x	x	UC 1:	Patientenbezogene Kooperationen der Gesundheitsakteure (z.B. Überweisungen)	x	x				x	x	
		x		x		x		x	UC 2:	Casemanagement: Koordination von Pflegebedürftigkeit		x				x		
		x							UC 3:	Institutionen (z.B. Praxen) suchen				x		x	x	
									UC 4:	Telemedizin (z.B. Online-Sprechstunde)				x		x		
	x	x		x					UC 5:	Termine online vereinbaren						x		
x	x	x	x				x	x	UC 6:	Versendung von eBefunde/Dokumente an verschiedene Gesundheitsakteure								
	x			x				x	UC 7:	elektronisches Entlassmanagement				x			x	
									UC 8:	Notfallsignal (Benachrichtigung einer Vertrauensperson)				x				
Organisation																		
x	x	x	x				x		UC 9:	Notfalldatensatz			x	x		x	x	
x	x	x		x				x	UC 10:	Kalender und/oder Erinnerungsmodul					x	x	x	
x	x	x	x	x			x	x	UC 11:	Dokumentenverwaltung (z.B. Arztbrief, Röntgenaufnahmen, Patientenverfügung, Rechnungen)	x	x	x	x	x	x	x	x
x	x	x		x		x	x	x	UC 12:	Verwaltung des Medikationsplan	x		x	x	x	x	x	x
x	x								UC 13:	Automatische Ausstellung von Folgerezepten						x	x	
x	x	x					x		UC 14:	Wechselwirkungscheck						x	x	
x	x	x		x					UC 15:	Impfungen verwalten (Impfpass)	x				x	x	x	x
		x							UC 16:	Implantatsausweis						x		
		x							UC 17:	Mutterpass	x					x	x	
Monitoring																		
						x			UC 18:	Telemonitoring (Familienmitglieder o.Ä.)			x					x
x	x	x		x		x		x	UC 19:	Prävention / Gesundheitscheck				x			x	
x	x	x	x	x			x	x	UC 20:	Vitalparameter dokumentieren / Patientenselbstdokumentation	x		x	x	x	x	x	x
x		x						x	UC 21:	Trainingsdaten erheben/Anbindung an Fitnesstracker/ -apps			x	x				
Forschung																		
	x	x	x		x				UC 22:	Datenfreigabe für klinische Studien				x				
Verwaltung und System																		
									UC 23:	Integration bestehender Leistungshistorie der Krankenkasse					x		x	
	x		x	x				x	UC 24:	Schnittstelle (z.B. KV-Connect, auch in Planung)		x		x	x		x	
		x							UC 25:	Mobile App verfügbar	x			x		x	x	x

3 Das Internet of Health als Ausprägung des Internet of Everything

3.1 Vier Bereiche des IoE im Gesundheitswesen

Im Gegensatz zum IoT, welches lediglich aus *Dingen* besteht, basiert das IoE zusätzlich auf den Säulen *Personen*, *Prozessen* und *Daten* [Mi15]. Evans (2012) stellt heraus, dass der Wert des IoE in der intelligenten Vernetzung dieser vier Säulen besteht. Der Vernetzungsgrad umfasst beim IoE machine to machine (M2M), people to machine (P2M) und people to people (P2P) Kommunikation, wie Abb. 1 veranschaulicht [Ev12]. Im Kern des IoE stehen Prozesse, die durch die zunehmende Vernetzung unterstützt werden.

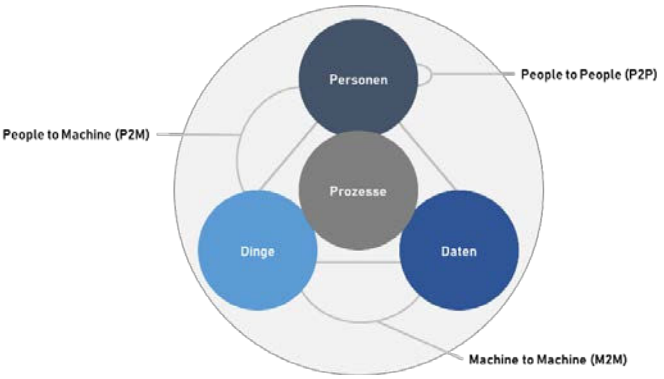


Abb. 1: Vier Bestandteile des Internet of Everything angelehnt an [Ev12]

Mit diesen vier Säulen ist das IoE in besonderem Maße auch für das Gesundheitswesen geeignet [BEE18], wie im Folgenden herausgearbeitet wird:

Personen. Die Gesundheitsversorgung ist durch das Zusammenwirken zahlreicher unterschiedlicher Akteure geprägt, die in drei Gruppen eingeteilt werden können [BFT19], [STS05]. Primäre Stakeholder sind Schlüsselnutzer, welche direkt an der Gesundheitsversorgung beteiligt sind. Dazu gehören Ärzte, Pfleger, Therapeuten, Apotheker, Krankenhäuser, Labore, Pflegedienste sowie Patienten. Zu den sekundären Stakeholdern zählen Versicherungen, Angehörige und Familienmitglieder sowie Arbeitgeber. Tertiäre Stakeholder umfassen die Politik, Gesellschaft, Forschungsinstitute, öffentliche Behörden und die Unternehmen aus der Gesundheitsbranche.

Prozesse. Solange ein Patient sich innerhalb einer Institution befindet, wird er in der Regel durch die Prozesse geführt und begleitet. Verlässt er hingegen die Institution, ist der Patient häufig auf sich allein gestellt. Der nächste behandelnde Akteur hat dann oft keinen Zugang zu bisherigen Diagnosen oder kennt weitere Erkrankungen des Patienten nicht [STS05]. Abstimmungen unter den Akteuren und Institutionen finden trotz zunehmender Digitalisierung häufig per Telefon statt. Diese Medienbrüche bei dem Übergang von verschiedenen Institutionen verursachen zum einen erhebliche Kosten durch erhöhten Koordinationsaufwand und führen zudem zu Qualitätsverlusten in der Versorgung. Insbesondere bei der Verschreibung von Medikamenten ist darauf zu achten, dass neue Wirkstoffe nicht mit bereits eingenommenen Medikamenten reagieren und unerwünschte Nebenwirkungen hervorrufen [NDW06]. Diese und zahlreiche andere Prozesse könnten durch eine nahtlose Vernetzung vereinfacht und weniger fehleranfällig gestaltet werden.

Daten. Im Rahmen der Gesundheitsversorgung entsteht eine Vielzahl an Daten in unterschiedlichen Formaten. Häufig können diese Daten aufgrund ihrer Vielseitigkeit nicht miteinander kombiniert werden. Dabei kann gerade diese Kombination wertvolle Erkenntnisse über die Gesundheitsversorgung hervorbringen [HB17]. Wenn zum Beispiel ein neues Medikament verschrieben wurde und eine dauerhafte Veränderung von Puls- oder Blutdruckwerten gemessen wird, könnte daraus eine Unverträglichkeit oder Wech-

selwirkungen mit anderen Medikamenten identifiziert werden. Darüber hinaus kann die systematische Auswertung von Gesundheitsdaten die Prävention verbessern, indem Krankheiten frühzeitig diagnostiziert werden [Ho11], [Mc15].

Dinge. Ebenso wie die Daten können auch die Dinge des IoE sehr vielseitig sein. In Bezug auf das Gesundheitswesen bedeutet dies, dass Nutzer ihre Gesundheitsdaten bestenfalls auf allen gewünschten Endgeräten abrufen können. Gleichzeitig nehmen die unterschiedlichen Dinge auch neue Daten auf. Mobile Geräte wie Tablet, Smartphone sind alltägliche Begleiter und sammeln zahlreiche Daten, die im Gesundheitskontext wertvoll sein können. Ebenso zeichnen Fitnesstracker unser Bewegungsprofil auf und messen unterschiedliche Vitalparameter. Wenn diese Geräte intelligent vernetzt werden, können daraus Informationen über den Gesundheitsstatus abgeleitet werden. Auch digitale Implantate sind ein Beispiel des IoE im Gesundheitswesen. Herzschrittmacher, die digital überwacht werden, oder implantierte Sensoren zur Blutzuckermessung können das Leben von chronisch erkrankten Patienten erheblich vereinfachen [Ca17], [Mo17]. Digitale Medikamentendosen können dem Patienten zu jeder Tageszeit die richtigen Medikamente freigeben und automatische Nachbestellungen auslösen. Dies vereinfacht besonders das Leben für ältere Menschen, die ihr Medikamentenmanagement nicht mehr selbstständig bewerkstelligen können. Für diese Personen können auch Smart Home oder Ambient-Assisted-Living-Technologien eine gute Möglichkeit sein, lange im eigenen Zuhause Wohnen zu können [Be17]. Technologien wie Sturzsensoren oder Telemedizinanwendungen helfen, die Gesundheits- und Pflegeversorgung zu verbessern.

Zusammenfassend lässt sich festhalten, dass das IoE das Potenzial hat, die Gesundheitsversorgung grundlegend zu verändern. Mit voranschreitender Vernetzung der Personen, Prozesse, Daten und Dinge kann so ein IoH als Ausprägung des IoE entstehen. Hierfür wird jedoch eine zentrale Plattform benötigt, die alle Bestandteile des IoH miteinander vernetzt. An diese Plattform werden zahlreiche Anforderungen gestellt. Neben einem benutzerfreundlichen Design, müssen zahlreiche Schnittstellen zu den Informationssystemen der Akteure geschaffen werden. Eine besondere Bedeutung kommt aufgrund der sensiblen Informationen dem Datenschutz und der Datensicherheit zu. Eine patienten- bzw. nutzerzentrierte eGA kann diesen vielseitigen Anforderungen gerecht werden und somit als Vernetzungsinstrument im IoH eingesetzt werden.

3.2 Die elektronische Gesundheitsakte als Vernetzungsinstrument im IoH

Bisher verwenden die einzelnen Akteure im Gesundheitswesen meist eigene Systeme, was dazu führt, dass die Gesundheitsversorgung nicht auf einem einheitlichen Datenbestand beruht. Die eGA bietet die Möglichkeit, die Systeme des Patienten und der verschiedenen Gesundheitsdienstleister wie u. a. Ärzte und Apotheker miteinander zu vernetzen [HBS17]. Mit einer tiefen Verankerung der eGA in die Prozesse im Gesundheitswesen dient sie als Vernetzungsinstrument im IoH, da sie die Personen, Daten, Dinge und Prozesse intelligent miteinander verbindet. Abb. 2 veranschaulicht das Potenzial der eGA als Vernetzungsinstrument im IoH. Die Verbindungen zwischen den Ele-

menten verdeutlichen, dass eine Plattform zur Informationsvermittlung zwischen den Bestandteilen unerlässlich ist. Diese Funktion kann die patienten- bzw. nutzergeführte eGA nach aktuellem Stand bestmöglich wahrnehmen, da beim Patienten alle Schnittstellen, die an der Gesundheitsversorgung beteiligt sind, zusammenlaufen. Der Patient kann beliebig viele *Dinge* wie z. B. Fitnesstracker, *Personen* wie Ärzte oder Angehörige, *Daten* zu seiner Behandlungshistorie und *Prozesse* wie Medikamentenbestellungen an die personalisierte eGA anbinden. Darüber hinaus ist es notwendig, neben der Vernetzung im Gesundheitswesen auch Schnittstellen zu anderen Bereichen zu bieten. Bspw. können auch aus dem Bereich Smart Home oder Mobilität wichtige Informationen abgeleitet werden, die für die Pflege notwendig sind [Be17]. Eine Krankenfahrt kann durch das Terminvereinbarungssystem des Arztes automatisiert gebucht werden oder im Falle eines Sturzes in der Wohnung kann für Angehörige oder Notfalldienste die Tür geöffnet werden. Durch die intelligente Vernetzung kann das IoH die Gesundheits- und Pflegeversorgung entscheidend verbessern und vereinfachen.

4 Internet of Health aus Sicht der Leistungserbringer

Um das Potenzial der eGA als Vernetzungsinstrument im IoH mit Vertretern aus der Praxis zu diskutieren, wurden Ende 2018 neun Experteninterviews mit verschiedenen Gesundheitsakteuren geführt (siehe Tab. 2) [GL10]. Gleichzeitig wurde diskutiert, welche Herausforderungen für einen flächendeckenden Einsatz einer eGA bestehen (FF3).

Tab. 2: Übersicht der befragten Experten

Nr.	Beschreibung	Dauer
E1	Apotheker	40 Min.
E2	Berater im Gesundheitswesen	31 Min.
E3	Apotheker	27 Min.
E4	Gründer und Geschäftsführer einer App für Apotheken	22 Min.
E5	Krankenpfleger und Case Manager	24 Min.
E6	Krankenhausapothekerin und Beraterin im Gesundheitswesen	36 Min.
E7	Geschäftsführer einer digitalen Patientenakte	24 Min.
E8	Krankenschwester	22 Min.
E9	Arzt (Orthopäde und Unfallchirurg)	44 Min.

In der voranschreitenden Vernetzung der Personen, Prozesse, Daten und Dinge durch die eGA sehen die befragten Experten großes Potenzial. Insbesondere die Sicherstellung der zeitnahen **Verfügbarkeit und Vollständigkeit der Gesundheitsdaten** für alle beteiligten Akteure wird als großer Vorteil genannt [E1, E2, E5, E9], [Am16]. So werden Therapiekonzepte und -abläufe für alle Akteure **transparenter**, sodass **Doppelverordnungen und Behandlungsfehler verhindert** werden können [E2, E6, E9]. Auch Wechselwirkungen können durch ein verbessertes Medikationsmanagement ausgeschlossen werden. Das ist vor allem bei Patienten mit aufwendiger Medikation, langer Krankheitsge-

schichte oder fehlender Orientierung, z. B. in Notfällen oder bei dementen Patienten, von besonderer Relevanz [E1, E3, E8]. Der flächendeckende Einsatz einer eGA unterstützt die Kommunikation und den Informationsfluss in allen drei Bereichen medizinischen Handelns (Diagnostik, Therapie und Nachsorge) und ermöglicht es, Versorgungs- und Informationslücken effizient zu schließen [E4, E7]. Der **effiziente Austausch** zwischen den Gesundheitsakteuren **vereinfacht die Kommunikation** zwischen den Beteiligten und dient dazu, weitere Akteure wie Pflegedienste, Pflegeheime, Apotheken und Sanitätshäuser einzubinden und zu unterstützen [E6], [STS15], [HB17]. Dies vereinfacht bei Diagnosen bspw. das Einholen von Zweitmeinungen und unterstützt die Therapieüberwachung und Nachsorge, indem die Weitergabe relevanter Dokumente sichergestellt wird [E3, E7]. Im Idealfall verfügen die Gesundheitsakteuren über alle relevanten Informationen, d. h. „*die Information kommt ‚mit' oder vor dem Patienten, sodass die Versorgung des Patienten vorbereitet ist, wenn er zum Versorger gelangt*" [E5].

Einen besonders hohen Mehrwert sehen die Experten in einer verbesserten Arzt-Patienten-Beziehung: während Patienten laut dem interviewten Berater im Gesundheitswesen bisher „*eher entmündigt*" sind, könnte die vom Patienten verwaltete eGA das **Patienten-Empowerment** steigern [E2, E7], [Am16]. Durch die Einbeziehung der vom Patienten geführten eGA können die Patienten selbst die Therapie unterstützen, indem sie bspw. dem behandelnden Arzt ergänzende Informationen wie Symptome oder Vitalwerte über eine Tagebuchfunktion zur Verfügung stellen, denn „*erst in dem Einbeziehen der Akte in die konkrete Therapie entfaltet sich der Nutzen*" [E7]. Neben der verbesserten Therapietreue und Kommunikationsmöglichkeiten sehen die Experten Potenziale für **Zeit- und Kosteneinsparungen**. So führen bspw. eine Abschaffung der papierbasierten Dokumentation [E1] und die Reduzierung von Ausdrucken [E2] zur Senkung der Druck- und Lagerhaltungskosten. Verkürzte Laufwege sowie das Vermeiden von unnötigen Arbeitsschritten und Doppeluntersuchungen beschleunigen die Prozesse woraus eine Effizienzsteigerung der Leistungserbringer resultiert [E3, E8, E9], [STS05].

Trotz des großen Potenzials des IoH durch die eGA äußerten die Experten vor allem Bedenken hinsichtlich des **Datenschutzes und der Datensicherheit**. Neben der Herausforderung der Zugriffsrechtsregelungen bestehe die Gefahr, dass sensible Daten durch unerlaubte Zugriffe missbraucht werden [E2, E4, E8, E9], [AMT19]. Insbesondere Einrichtungen der Privatwirtschaft und Krankenversicherungen hätten ein finanzielles Interesse an den Daten und könnten diese für Kosteneinsparungen, Optimierung der Preispolitik und als Medium für Qualitätsuntersuchungen nutzen [E1, E2, E5, E6, E9]. Neue Möglichkeiten zur Absicherung sensibler Gesundheitsdaten ergeben sich durch die Blockchain-Technologie [BFT19]. Die Politik sei zudem in der Verantwortung, **gesetzliche Rahmenbedingungen** zu schaffen, da eGA-Anbieter die größten Hürden in Datenschutzbestimmungen und Haftungsfragen sehen [E3, E6], [Ha17]. Die eGA könnte hingegen auch genutzt werden, um durch die Autorisierung des Patienten Klarheit in Datenschutzfragen zu schaffen [E1]. Zur Garantie des sicheren Einsatzes der eGA bedarf es außerdem einer bundesweiten Nachschulung aller Ärzte und Nutzer der eGA [E8, E9].

Die Krankenhausapothekerin [E6] stellt darüber hinaus die Gewährleistung der Vollständigkeit der eGA in Frage. Wenn die Datenhoheit beim Patienten liegt, bestehe die Gefahr einer **Selektion des Informationsflusses**, da der Patient jeden Zugriff aktiv freigeben müsse und mögliche Daten vorenthalten könne, deren Relevanz er nicht richtig einschätzen kann [E6], [Mo17]. Hinzu kommen Zweifel, ob Patienten die eGA ab einem gewissen Alter noch selbst verwalten können [E8]. Laut dem interviewten Anbieter einer digitalen Patientenakte könne sich das auf das **Vertrauen** der Ärzte in die eGA auswirken: „*wenn die Akte dem Patienten gehört, vertrauen die Ärzte der Akte nicht, Ärzte wollen Daten von anderen Ärzten sehen*“ [E7]. Zu den bisherigen Herausforderungen eines flächendeckenden Einsatzes der eGA zählt insbesondere auch die Einstellung der Ärzte, welche sich aufgrund der Ineffizienz und Mehrarbeit häufig von bisherigen digitalen Erneuerungen und Produkten verschließen. „*Ärzte sind es gewohnt, Dinge am Patientenbett zu notieren und zu besprechen und erst hinterher im Arztzimmer zu digitalisieren*“ [E6]. Laut den Experten bedarf es einem Umdenken der Leistungserbringer. Abschießend fordern die Experten die Einführung einheitlicher Standards im Umgang mit eGAs, um die bundesweite bzw. EU-weite Interoperabilität zu gewährleisten und Fehlinvestitionen der Akteure vorzubeugen [E1, E3, E4, E7]. Letztendlich bestehen die größten Herausforderungen laut den Experten immer noch in der technischen Umsetzung und der fehlenden **Infrastruktur** [HB17]. Es müssten Server, Kartenlesegeräte und eine verschlüsselte Verbindung etabliert werden [E4, E5, E6, E9]. Bislang gibt es viele verschiedene Systeme mit unterschiedlichen Dateiformaten, die und eine Kopplung an bestehende Arztinformationssysteme erschweren [E1, E6, E7]. Die Schaffung einheitlicher **Standards** seitens der Politik wird daher von allen Experten als zwingend erforderlich gesehen. Das im Rahmen dieser Untersuchung identifizierte Potenzial der eGA als Vernetzungsinstrument im IoH (FF2) wird zusammen mit den zentralen Herausforderungen (FF3) in Abb. 2 zusammengefasst.

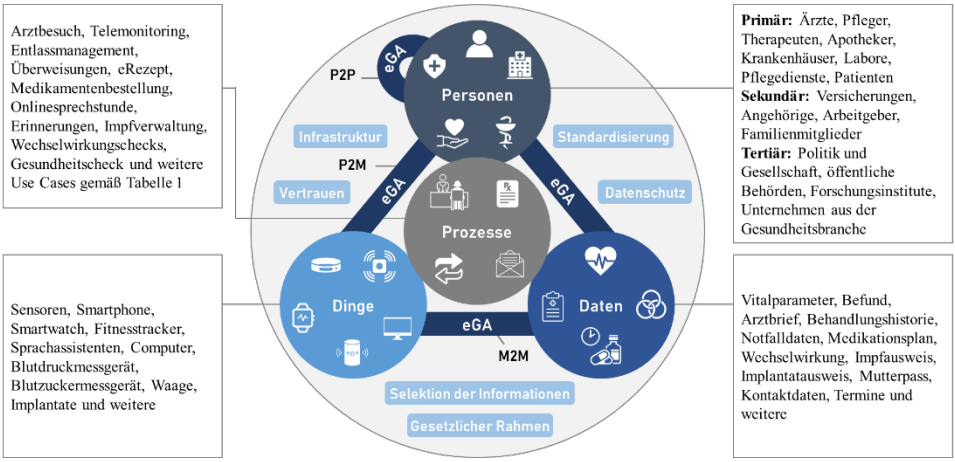


Abb. 2: Bedeutung der elektronischen Gesundheitsakte im Internet of Health

5 Diskussion

Die systematische Untersuchung der eGA und die Diskussion mit Akteuren der Gesundheitsversorgung haben gezeigt, dass die eGA ein notwendiges Vernetzungsinstrument ist, um die vielfältigen Personen, Prozesse, Daten, und Dinge im IoH zu vernetzen. Mit dem flächendeckenden Einsatz der eGA haben alle Akteure, die an der Gesundheitsversorgung beteiligt sind, die Gelegenheit, auf einen einheitlichen Datenbestand zuzugreifen. Mit der Verbreitung der eGA wird eine starke Verankerung in die Prozesse (z. B. Überleitungen zum Facharzt, eRezept) einhergehen. Durch die Einbindung unterschiedlicher Dinge und Daten kann die Qualität der Gesundheits-versorgung erhöht werden, weil für Diagnosen unterschiedliche Quellen einbezogen werden. Insbesondere im Bereich Smart Home ergeben sich große Potenziale für die Versorgung von pflegebedürftigen Menschen im eigenen Zuhause [Be17]. Da es sich bei Gesundheitsdienstleistungen immer um sensible Daten handelt, muss bei der Umsetzung darauf geachtet werden, dass die Infrastruktur sicher ist und genügend Maßnahmen bzgl. des Datenschutzes bspw. in Form von 2-Faktor-Authentifizierung getroffen werden. Dies gestaltet sich besonders schwierig, da aufgrund der Skalierung auf Cloud-Dienste zurückgegriffen werden muss. Jedoch haben viele Anwender bei Cloud-Diensten Bedenken bzgl. der Sicherheit und Privatsphäre [AMT19]. Aus diesem Grund muss bei der Gestaltung der eGA darauf geachtet werden, dass die Nutzer über die Verwendung sowie der Sicherheit ihrer Daten informiert werden und sie jederzeit die Möglichkeit haben, auf die Daten zuzugreifen und die Zugriffsrechte auf diese zu kontrollieren. Eine gleichermaßen vertrauensfördernde und effiziente Maßnahme wäre die Einführung von verbindlichen Standards [FB18]. Einerseits würde die Anbindung an die Systeme der Dienstleister erleichtert werden, andererseits hätten die Akteure Planungssicherheit, welcher Standard flächendeckend eingesetzt wird.

Es hat sich herausgestellt, dass die Anwendungsfälle die Prozesse im Gesundheitswesen in vielen Bereichen vereinfachen können, die Umsetzung durch die staatlichen Initiativen jedoch sehr langsam verläuft. Die meisten Experten schätzen, dass der zeitnahe Einsatz der eGA durch die Patienten vorangetrieben werden muss. Als primäre Nutzer müssen die Patienten eine eGA von privaten Anbietern beziehen und ihre Gesundheitsdaten selbstständig darin speichern. Abhängig von der eGA können die Patienten den Leistungserbringern Nutzungsrechte geben. So können auch diese Akteure Informationen speichern. Der vorgestellte Ansatz einer eGA ist gekennzeichnet durch zahlreiche Schnittstellen zu anderen IoT-Systemen. In diesem Kontext ergeben sich Fragen bezüglich der Datenhoheit sowie der technischen Realisierung. Nach den Expertenmeinungen werden sich wenige eGA Anbieter durchsetzen können und eine kritische Masse erreichen, sodass die Systeme der Dienstleister wie u. a. Krankenhaus- und Arztinformationssysteme ihre Schnittstellen anpassen, um Daten direkt auszutauschen [HBS17]. Auf diese Weise ist sowohl von der Seite der Patienten als auch aus Sicht der Leistungserbringer weniger Aufwand nötig, damit die Daten kontinuierlich gepflegt sind und somit für eine fundierte Versorgung genutzt werden können. So wird nicht nur der §630g BGB erfüllt, durch den Patienten ein Recht darauf haben, Einsicht auf ihre ePA zu erhalten,

sondern es besteht die Möglichkeit, dass die Patienten ihre Daten aus der ePA direkt zur Verfügung gestellt bekommen. Die Experten sind sich jedoch einig, dass es auch bei dem flächendeckenden Einsatz einer eGA weiterhin ePAs genutzt werden.

Neben den direkten Leistungserbringern haben auch Krankenkassen ein Interesse daran, dass ihre Versicherten eine eGA nutzen. Durch die bessere Informationsversorgung, die durch eine eGA erreicht werden kann, können die Krankenkassen Geld einsparen, da Untersuchungen nicht doppelt vorgenommen werden müssen. Dieses Angebot kann dazu führen, dass viele Versicherte sich dazu entscheiden, eine eGA auszuprobieren und bei einem erkenntlichen Mehrwert langfristig zu adoptieren. Dafür ist es jedoch auch notwendig, dass die eGAs für die Anwender einfach nutzbar sind. Dies kann erreicht werden, indem die Anbieter ihre Patientenakten an dem Nutzungsverhalten der Anwender ausrichten. Dabei muss neben der Nutzbarkeit darauf geachtet werden, dass die Prozesse im Gesundheitswesen durch die eGA unterstützt werden.

6 Fazit

Ausgangspunkt des vorliegenden Beitrags war die Frage, wie die eGA als Vernetzungsinstrument im IoH eingesetzt werden kann. Hierfür wurden basierend auf den Ergebnissen einer Literaturrecherche und der Überprüfung von sechs aktuell umgesetzten eGAs 25 Anwendungsfälle identifiziert, in denen die eGA bei der Gesundheitsversorgung unterstützen kann (FF1). Durch die Untersuchung des aktuellen Standes der Wissenschaft in Kombination mit einer Analyse der aktuell umgesetzten Lösungen wird ein ganzheitliches Bild für den Einsatz von eGAs im deutschen Gesundheitswesen geboten. Dies bietet weiteren Forschungsarbeiten eine gute Grundlage sowie Gesundheitsdiensteistern und Patienten einen Überblick über die Möglichkeiten der Gesundheitsakten. Den Anbietern von eGAs ermöglichen die identifizierten Anwendungsfälle eine Übersicht um welche Komponenten ihre Produkte erweitert werden könnten, um eine ganzheitliche Abdeckung zu erreichen. Anschließend wurde herausgearbeitet, inwiefern die eGA als Vernetzungsinstrument im IoH eingesetzt werden kann (FF2). Die Untersuchung hat ergeben, dass eine zentrale Vernetzungsplattform unerlässlich für die Etablierung eines IoH ist. Die eGA hat das Potenzial, den vielseitigen Anforderungen an diese Plattform gerecht zu werden. In der anschließenden Evaluierung durch neun Experteninterviews wurden die Anwendungsmöglichkeiten der eGA mit verschiedenen Akteuren im Gesundheitswesen diskutiert und Herausforderungen für eine flächendeckende Verbreitung identifiziert. Limitierend ist anzumerken, dass nicht alle relevanten Stakeholder in der Untersuchung berücksichtigt werden konnten. Insbesondere Patienten wurden nicht in die Erhebung einbezogen, da für valide Ergebnisse eine quantitative Erhebung notwendig ist. Darüber hinaus konnte nur eine Auswahl deutscher eGA Anbieter abgedeckt werden. Weiterer Forschungsbedarf ergibt sich zudem bezüglich der Finanzierung der eGA sowie in einer Kosten-/Nutzen-Analyse. Dennoch stellt diese Untersuchung eine systematische Aufarbeitung der Potenziale der eGA für die Gesundheitsversorgung der Zukunft dar.

7 Danksagung

Diese Publikation ist im Rahmen der Forschungsprojekte **Dorfgemeinschaft 2.0** (BMBF/ Projektträger VDI/VDE-IT) und **Apotheke 2.0** (Bundesprogramm Ländliche Entwicklung, Bekanntmachung „Land.Digital“, BMEL/ PT BLE) entstanden.

Literatur

[Am05] Ambinder, E. P.: Electronic Health Records. Journal of Oncology Practice 1 (2005) 2, S. 57–63.

[Am16] Amelung, V.; Bertram, N.; Binder, S.; Chase, D. P.; Urbanski, D.: Die elektronische Patientenakte. In: Fundament einer effektiven und effizienten Gesundheitsversorgung. Stiftung Münch (Hrsg.), medhochzwei (2016).

[AMT19] Adelmeyer, M.; Meier, P.; Teuteberg, F.: Security and Privacy of Personal Health Records in Cloud Computing Environments–An Experimental Exploration of the Impact of Storage Solutions and Data Breaches. Internationale Tagung Wirtschaftsinformatik, 2019.

[Be17] Beinke, J. H.; Meier, P.; Nickenig, H.-P.; Teuteberg, F.: Smart Home Predictive Analytics. Informatik 2017.

[BEE18] Bauer, C.; Eickmeier, F.; Eckard, M.: E-Health: Datenschutz und Datensicherheit - Herausforderungen und Lösungen im IoT Zeitalter. (2018).

[BFT19] Beinke, J. H.; Fitte, C.; Teuteberg, F.: Towards a Stakeholder-oriented Blockchain-based Architecture for Electronic Health Records. Journal of Medical Internet Research (2019).

[Ca17] Cappon, G.; Acciaroli, G.; Vettoretti, M.; Facchinetti, A.; Sparacino, G.: Wearable continuous glucose monitoring sensors: A revolution in diabetes treatment. Electronics 6 (2017) 3, S. 65.

[Ev12] Evans, D.: How the Internet of Everything Will Change the World. Cisco Blog, 2012.

[FB18] Flaumenhaft, Y.; Ben-Assuli, O.: Personal health records, global policy and regulation review. Health Policy 122 (2018) 8, S. 815–826.

[Fo19] Forecare, 2019. URL: https://www.forcare.com/ (04. April 2019).

[GL10] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse. Springer, 2010.

[Ha17] Haas, P.: Elektronische Patientenakten. Berteslmann Stiftung, 2017.

[HB17] Heinze, O.; Bergh, B.: Persönliche einrichtungsübergreifende Gesundheits-und Patientenakten (PEPA) als zentrale Infrastrukturkomponente einer patientenzentrierten Gesundheitsversorgung. E-Health-Ökonomie. Springer, 2017, S. 847–858.

[HBS17] Heart, T.; Ben-Assuli, O.; Shabtai, I.: A review of PHR, EMR and EHR integration: A more personalized healthcare and public health policy. Health Policy and Technology

6 (2017) 1, S. 20–25.

[He19] HealthVault 2019. URL: https://international.healthvault.com (04. April 2019).

[Ho11] Hogan, T. P.; Wakefield, B.; Nazi, K. M.; Houston, T. K.; Weaver, F. M.: Promoting access through complementary eHealth technologies: recommendations for VA's Home Telehealth and personal health record programs. Journal of general internal medicine 26 (2011) 2, S. 628.

[HSN08] Häyrinen, K.; Saranto, K.; Nykänen, P.: Definition, structure, content, use and impacts of electronic health records: a review of the research literature. International journal of medical informatics 77 (2008) 5, S. 291–304.

[Kr18] Krüger-Brand, H. E.: Elektronische Gesundheitsakten: Erster Anbieter prescht vor, aerzteblatt.de, 2018.

[Mc15] McCowan, C.; Thomson, E.; Szmigielski, C. A.; Kalra, D.; Sullivan, F. M.; Prokosch, H.-U.; Dugas, M.; Ford, I.: Using Electronic Health Records to Support Clinical Trials: A Report on Stakeholder Engagement for EHR4CR. BioMed Research International (2015), S. 1–8.

[Mi15] Miraz, M. H.; Ali, M.; Excell, P. S.; Picking, R.: A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT). Internet Technologies and Applications - Proceedings of the 6th International Conference (2015), S. 219–224.

[Mo17] Monica, K.: 6 Use Cases for EHR Data Utilization in Public, Community Health, 2017. URL: https://ehrintelligence.com/news/6-use-cases-for-ehr-data-utilization-in-public-community-health (17. April 2019).

[NDW06] Neuhaus, J.; Deiters, W.; Wiedeler, M.: Mehrwertdienste im Umfeld der elektronischen Gesundheitskarte. Informatik-Spektrum 29 (2006) 5, S. 332–340.

[Pa19] PatientAssist, 2019. URL: http://www.patientassist.de (04. April 2019).

[Pr01] Prokosch, H.-U.: KAS, KIS, EKA, EPA, EGA, E-Health: Ein Plädoyer gegen die babylonische Begriffsverwirrung in der Medizinischen Informatik. Informatik, Biometrie und Epidemiologie in Medizin und Biologie 32 (2001), S. 371–382.

[RJ17] Rohleder, B.; Jedamzik, S.: Gesundheit 4.0. (2017).

[Ro18] Rodrigues, J. J. P. C.; Segundo, D. B. D. R.; Junqueira, H. A.; Sabino, M. H.; Prince, R. M.; Al-Muhtadi, J.; De Albuquerque, V. H. C.: Enabling technologies for the internet of health things. Ieee Access 6 (2018), S. 13129–13141.

[STS05] Schwarze, J.; Tessmann, S.; Sassenberg, C.: Eine modulare Gesundheitsakte als Antwort auf Kommunikationsprobleme im Gesundheitswesen. Wirtschaftsinformatik 47 (2005) 3, S. 187–195.

[TK19] Techniker Krankenkasse: TK-Safe startet, 2019. URL: https://www.tk.de/presse/themen/digitale-gesundheit/digitale-gesundheitsakte/tk-safe-2039872 (04. April 2019).

[Vi19] Vitabook, 2019. URL: http://www.vitabook.de (04. April 2019).

A Context Map as the Basis for a Microservice Architecture for the Connected Car Domain

Sebastian Abeck[1], Michael Schneider[2], Jan-Philip Quirmbach[3], Heiko Klarl[4], Christof Urbaczek[5] and Shkodran Zogaj

Abstract: In the near future cars will have two properties: They will be electrically powered and they will be connected to the Internet. Such cars will provide a huge amount of sensor data which can be accessed via web APIs in order to develop innovative connected car applications, such as traffic control, hazard warning, assisted or even autonomous driving. However, current software solutions in this field are mainly monoliths solving single problems in an isolated way. Therefore, we propose a systematic approach by which each single connected car application becomes part of a microservice architecture. This approach requires a sound and well-elaborated domain model from which the microservices' APIs and implementation of the applications can be systematically derived. The main contribution of this paper is a context map for the connected car domain. We demonstrate a structured software development approach with the example of a mobile application, the Electric Car Charger, by showing how this application is integrated into the context map and, thus, into a connected car microservice architecture.

Keywords: Connected car, microservice architecture, domain modeling, context map, bounded context, API

1 Introduction

Connected cars are in the center of innovative and complex mobility concepts for our society [Co+16]. Such mobility solutions, in which cars are only one means of transportation besides bus, train, bikes etc., requires the exchange of data between all involved transportation means (vehicle-to-vehicle) and the transportation infrastructure (vehicle-to-infrastructure) via the Internet. Therefore, the Internet of Things (IoT) aspect plays an important role in the field of integrated mobility solutions [DK+18]. Connected cars are one of these "things" of the Internet for which such new mobility services are offered. They can be seen as the drivers of IoT-based mobility solutions resulting from the economic power of the automotive industry. The necessary movement towards e-cars and their integration into an overall Internet-based mobility infrastructure lead to disruptive changes in this industrial domain. Besides the traditional automobile manufacturers offering cars as a product to their customers, new companies from the IT domain appear

[1] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, sebastian.abeck@kit.edu
[2] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, michael.schneider@kit.edu
[3] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, jan-philip.quirmbach@student.kit.edu
[4] xdi360, Munich, Germany, heiko.klarl@xdi360.com
[5] xdi360, Munich, Germany, christof.urbaczek@xdi360.com

doi:10.18420/inf2019_18

on stage. They perceive the cars as things of the Internet and provide connected car services. Examples of such services are shown in Fig. 1.

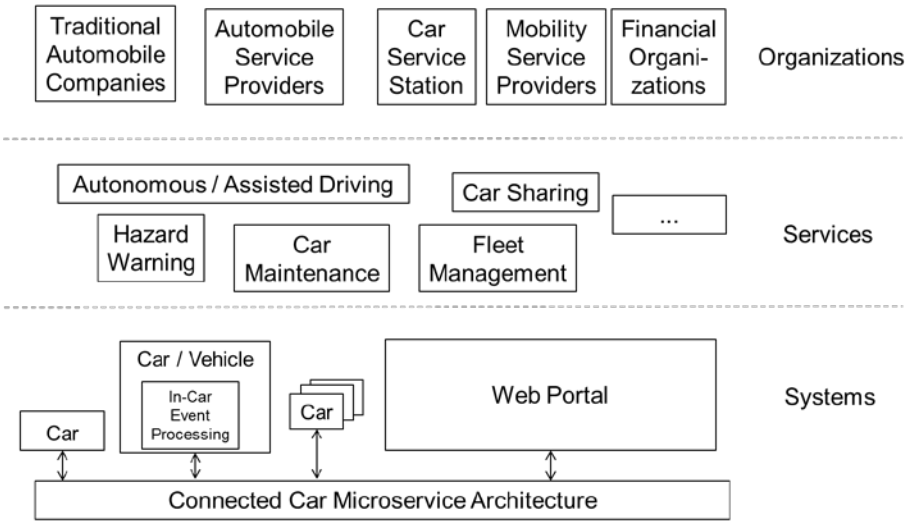


Figure 1: Examples of Connected Car Services

To be able to develop flexible and maintainable connected car solutions, a software architecture is needed which can be easily extended by new functional services. These services are provided by, and offered to, different organizations. We believe that a microservice architecture [Ne15] is an adequate concept to build a connected car system. This system consists of loosely coupled connected car services using other services via web APIs specified in a standardized language (e.g. OpenAPI). The microservice architecture is based on a domain model [Ev03] which prescribes the functional structure of the connected car domain.

Starting from the related work (Section 2), this paper elaborates a context map and its included bounded contexts of the connected car domain model (Section 3). The usage of the elaborated domain model artifacts is shown with the example of an Electric Car Charger (ECC) service (Section 4). The main advantage of our approach is the integration of the connected car service, in our case the ECC, into the overall connected car domain. The domain model and the derived microservice architecture provide the basis for all connected car services leading to a non-monolithic, loosely coupled connected car system (Section 5).

2 Related Work

Intensive work on digital technology and software engineering in the automotive sector started about the turn of the millennium [Br03]. The main competence of car companies traditionally lies in the field of mechanical and electrical engineering. In order to cope with the high complexity of automotive software, frameworks specific for the automotive domain were developed, such as the Volvo Cars Architecture Framework [PK+16] or the Automotive Architecture Framework [BG+09]. A characteristic of this work is the focus on the architecture of the software that is needed in a car. In [PK+16] the aspect of connected cars is covered in two so-called viewpoints, namely "connected cars and safety" and "security and privacy of connected cars". Although such automotive frameworks cover certain aspects of the automotive domain, they do not provide a domain model which is one of the main goals of this paper.

The Domain-Driven Design (DDD) [Ev03, Ve13] provides the conceptual foundation of our approach. As shown by [SH+18, HG+17], DDD can be applied in a structured software development process in order to derive a sound and comprehensible microservice architecture. A central part of the domain model is the so-called context map which is the result of DDD's strategic modeling. A context map is used to decompose the domain into subject-specific (especially not technically-driven) parts which are called bounded contexts. Since each bounded context is a candidate for a microservice [Ne15], the context map can be seen as a blueprint of the microservice architecture for the modeled domain. In [TH+18] a systematic approach to derive the bounded contexts in order to identify microservices is presented. The functional decomposition is carried out based on the requirements on the software system. A characteristic of this approach is given by the fact that a concrete software system, and not the domain, is in the focus. Therefore, a context map of the domain is not part of this approach.

Existing white papers from different companies (e.g. [KA+16, VA+14, DK12]) provide a fine-grained decomposition of a connected car's application landscape into different categories, such as navigation, vehicle management, or safety. This related work describes the domain in a more or less informal way. Nevertheless, for our work they provide a valuable practical input for the formal connected car domain model which we develop in the next Chapter 3 and apply to build a microservice-based application in Chapter 4.

3 Connected Car Categories

In the related work, different categories for the connected car domain are proposed which are illustrated in Figure 2. Vehicle management and driving management are directly related to the core functionality of a car. The category vehicle management is divided into the sub-categories remote control, diagnosis and maintenance; subcategories of driving management are driver assistance, parking, and refueling [KA+16].

Further, there exists a cross-cutting functionality safety and security. Safety and security need to be concerned by all other categories, most important for vehicle management and driving management, because critical functionalities need to be secure. For example, it should not be possible that one can remotely control a car without permission. Safety and security can be divided into further sub-categories, such as emergency and theft protection.

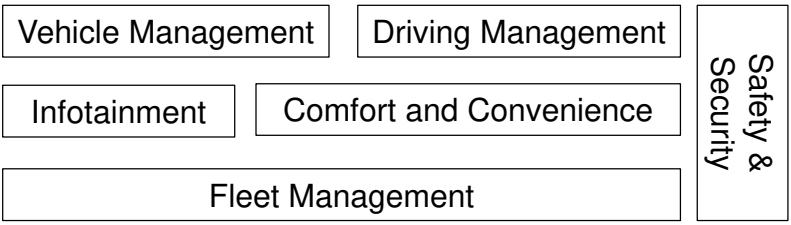


Figure 2: Existing Connected Car Categories

Infotainment as well as comfort and convenience provide less critical, but relevant functionalities. The infotainment category implies entertainment, information and smartphone integration and deals with streaming music and videos, interacting with social networks and providing news and weather information. Hand-free calls are an example of smartphone integration. Furthermore, information about the current traffic and navigation are also attributed to this category. Comfort and convenience are divided into the sub-categories well-being, interaction, and payment. Comfort and convenience include personalizing the vehicle, for example by pre-setting seats, temperature or ambience lighting.

In addition, issues referring to the governance of services across many cars deal with the category fleet management which consists of sub-categories like policies and optimization [DK12].

4 Formalization Based on UML and Domain-Driven Design

We use the existing proposals of a decomposition of the connected car domain to develop a formalized domain model based on the Domain-Driven Design (DDD, [Ev03]). This domain model serves as a design artifact from which we derive the microservice architecture for all connected car applications we are developing. The approach is not specific to the connected car domain since we use it also for other complex domains.

The context map displays the strategical relationships of a domain [HS+19, HG+17]. A context map consists of subdomains, bounded contexts and relationships between the bounded contexts. Following DDD, the bounded contexts are assigned to domain-specific subdomains, which further improve the overview of the domain. According to our approach, subdomains are modeled as a UML package which is extended by the stereotype <<subdomain>>.

A bounded context represents a candidate for a microservice which can be developed by an independent team [Ne15]. We formalize a bounded context as a packaging component which is annotated with the stereotype <<bounded context>>. Each bounded context contains tactical models like the relation view which describes the inner structure of this bounded context [SH+18].

Relationships between bounded contexts are formalized using UML associations extended by stereotypes corresponding to the context map relation. Depending on the type of relationship, the team communication between the bounded contexts is defined. [Ev14] provides several communication patterns for the relationships between bounded contexts. For example, the pattern <<conformist>> is a directed association between two bounded contexts. The consuming service has no influence on the offering service. Foreign bounded contexts are encapsulated by an <<anti-corruption layer>> (ACL). The ACL is formalized as a package which is part of the bounded context that uses the foreign bounded context.

5 Context Map for the Connected Car Domain

A decomposition of the connected car domain into subdomains and bounded contexts based on the formalization is derived. The context map, as shown in Figure 3 displays the result of the formalized connected car domain and suggests a separation of the different software services. For an easy overview and a better understanding, we put the main subdomains and bounded contexts in the center. Cross-section bounded contexts are placed on the right side of the context map diagram. Domain-enhancing bounded contexts that have a stronger user interaction are placed above the central area, and, finally, domain-supplementing bounded contexts, which express a more technical content, are located below the central area. Subdomains and bounded contexts that are close together are modeled in close proximity.

The context map is a design artifact of a structured software development process for microservice-based applications. Typically, CamelCase and PascalCase are used as a naming convention in such software development artifacts (e.g. VehicleManagement instead of vehicle management or Vehicle Management).

The category vehicle management offers a good starting point for the derivation of the subdomain VehicleManagement. We see these sub-categories as services for the vehicle management and therefore, bounded contexts for vehicles, sensor processing, remote control, diagnostics, and maintenance are established for this subdomain. The bounded context SensorProcessing processes the raw sensor data and provides semantically

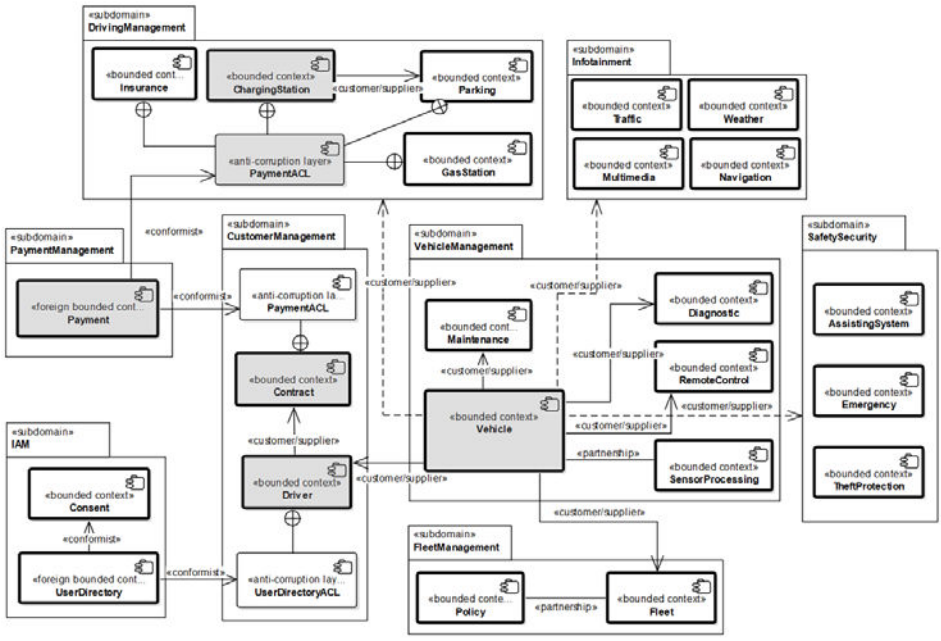


Figure 3: Proposed Context Map of the Connected Car Domain

enriched IoT data via an API. An example of how IoT platforms manage their sensor data with an IoT gateway and offer their sensor data by providing an API is given in [MK+17]. Functionalities for one remote controlling the vehicle are offered by the bounded context RemoteControl. The bounded context Diagnostic includes aspects like driving behavior analysis and telemetry data transmission. The bounded context Maintenance uses diagnostic data to perform predictive maintenance. If necessary, remote maintenance is handled by this bounded context. In addition, information from the bounded context Vehicle can be used to determine the functions supported by the vehicle. The bounded context Vehicle is one of the most important ones, because it offers the data base for many other bounded contexts.

In addition to the vehicle management, a customer management is required. This subdomain is needed in a connected car domain and the derived microservice architecture, even though no such category exists. Therefore, we added the subdomain CustomerManagement. This subdomain manages the data of the customer. For example, only the owner (or privileged users) of the car should be allowed to use the remote-control service for locking and unlocking the vehicle. The user-specific information is handled by the bounded context Driver. Since the customer is bound to contracts, we added a bounded context Contract. The customer data could be provided by a foreign identity and access management (IAM) system. The bounded context Driver uses the foreign bounded context UserDirectory. The UserDirectoryACL provides an additional

layer and handles the transformation of the external and internal data for the bounded context Driver. Thus, the bounded context Driver can use its own data representation.

In our context map, the category driving management results in a subdomain DrivingManagement. An assisting system for parking helps the driver to simplify the parking process, whereas a parking system supports the driver to find free parking lots. In addition, services offering information about gas stations are part of this subdomain. Thus, we derived the bounded contexts Parking and GasStation which provide the necessary information. A connection to an external payment service could simplify or fully automate the payment process.

Further, there is the infotainment category which implies entertainment, information and smartphone integration. These services are outsourced into independent bounded contexts as Traffic, Multimedia, Weather, and Navigation, in order to be able to adequately handle the underlying domain logic. This is necessary to guarantee the understanding and uniform representation of the information. For example, multimedia can also be separated into a bounded context, which takes over the connection to third parties and ensures uniform formats for video, image, and audio.

One of the most relevant subdomains in the context map of the connected car domain is SafetySecurity. We established a bounded context for each of the subcategories, since each of these can be encapsulated as a separate service: The bounded context TheftProtection may offer an alarm (locally and on the smartphone), as well as the tracking of the vehicle on the smartphone and automatic associated damage reports. In case of a technical defect or an accident, the bounded context Emergency can process the data. Through the connection to the bounded context Vehicle, relevant vehicle-specific data can be automatically retrieved and transmitted for the intervention teams.

The classification of the category comfort into the microservice architecture is not straightforward because the subdomain Comfort does not fit the connected car domain. However, several subdomains can be derived from this category. A payment provider is required to process the payment. Therefore, a new bounded context Payment is derived and placed into the subdomain PaymentManagement. Further, the category comfort contains the automatically pre-setting of the seat position for the driver. For this reason, the bounded context Driver manages the driver together with their data and personalized vehicle settings which could be used for a car sharing application.

The fleet management is particularly relevant for car sharing or for company fleets. A distinction between analysis, optimization, and guidelines is important for this area. The owner of the vehicle is a company, while the driver is an employee. Based on this, the functionalities and authorizations differ in the context of fleet management; the company is given access to data such as locations (logistics/just-in-time) and fleet consumption. These issues are captured in the subdomain FleetManagement. The bounded context Fleet is responsible for cross-fleet analysis, while the bounded context Policy can be used to define certain rules that must be observed by the vehicles in the fleet.

6 Case Study: Electric Car Charger Application

The connected car context map we have developed in the last chapter builds the fundament for all applications of this domain. One such application is the Electric Car Charger (ECC) which we informally describe in Section 6.1. In the following Section 6.2, we show how ECC fits into the connected car context map and we illustrate the development process with the example of the ECC application. We summarize the benefits of the context map's use for software development in Section 6.3.

6.1 ECC Domain Objects and Relationships

The ECC application implements a software solution for charging stations for electric cars. This application allows the user to search for charging stations displayed on a map view. Furthermore, it is possible to filter these stations based on several attributes, e.g. by plug type. The ECC also enables a monitoring function during the charging process to obtain further information during the charging.

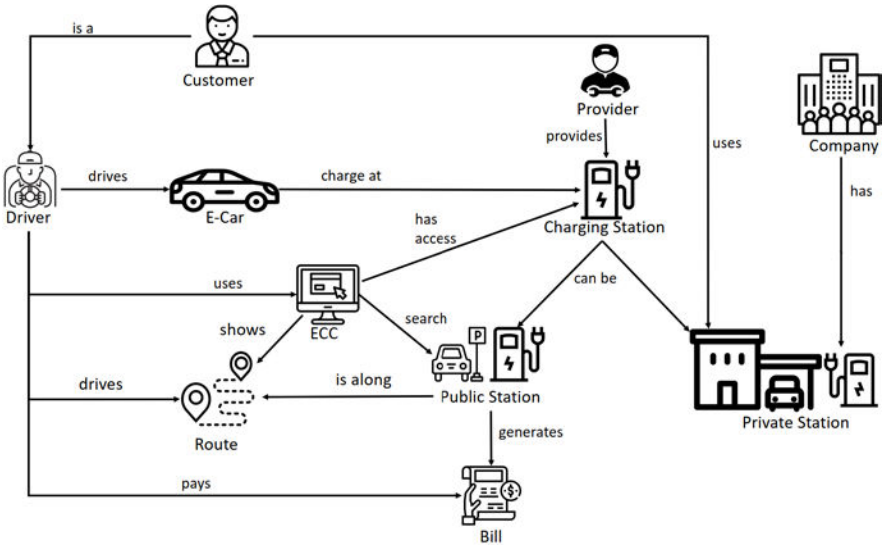


Figure 4: Domain Sketch of the ECC Application

Figure 4 shows a sketch of all relevant subjects and objects and their relations. The sketch provides an overview of the application-related part of the domain. The central element of the sketch is the charging station. A charging station can be a public station which could be installed, for example, at a parking lot or a private station which can be installed from a company at its private property. A customer is also the driver of the e-car. A driver can use the ECC application to (i) get information about the charging station, (ii) monitor the charging process, (iii) search a public station, and (iv) show all

public stations along a certain route the driver wants to take. An e-car can charge its battery at a charging station which a provider provides. When an e-car is charged at a public station the station generates a bill.

6.2 Use of the Context Map

The ECC context map was developed with the help of the overall context map of the introduced connected car domain. The proposed context map in Figure 3 shows the placement of the ECC by dyeing the relevant ECC objects in grey. The ECC-relevant parts of the context map were identified as follows: The main part for the ECC is the bounded context ChargingStation in which the ECC application was developed. The bounded context Vehicle is required to access the information concerning the battery of the vehicle. Due to the fact that the vehicle is related to the bounded context Driver, it is possible to get information about the driver. The driver also provides the contracts of the driver through the bounded context Contract. The contract for the usage of a private charging station is stored in this bounded context. The relationship to the bounded context Payment is required for the payment of the charging process.

The development process that is used during the development of the ECC application is based on Behavior-Driven Development (BDD) [SM15] and Domain-Driven Design (DDD) [Ev03]. Figure 5 shows how the context map is related to software development artifacts. During the analysis phase, the required functionalities are written in Gherkin features which are the central BDD artifact. The advantage of Gherkin is that the features can not only be written in a human-readable way, but also be executed and tested. Each Gherkin feature belongs to one bounded context, which is also a candidate for a microservice. A bounded context usually consists of several features.

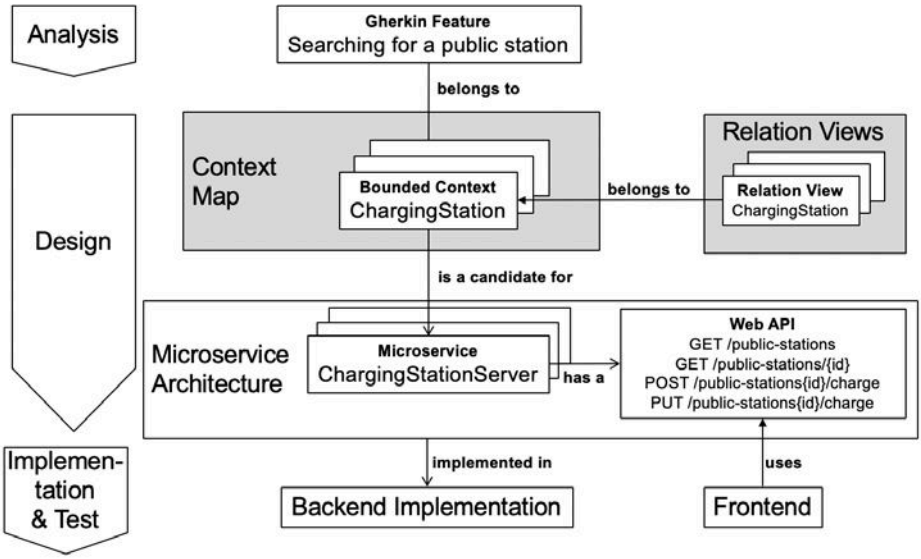


Figure 5: Context Map and related Software Development Artifacts

One feature of the ECC is searching for a public station (see Figure 5). This feature should display all public stations on a map view. During the design phase, the context map for the ECC application was designed. The ECC application was realized in the bounded context ChargingStation. For the technical interface of the resulting microservice, a web API based on the architectural style REST (REpresentational State Transfer, [Fi00]) was designed which offers the required functionality of the ECC. Figure 6 shows an excerpt of the Swagger UI for the request GET/public-stations/{id}.

The implementation of the backend and frontend was split and developed by two teams that could work almost independently of each other. The frontend team implemented the graphical user interface for the ECC, whereas the backend team implemented the required functionality in the backend and exposed it via the web API. The data of the public stations can be accessed via one of the web API methods, in particular the GET /public-stations method.

GET /charging-stations/public-stations/{id} Finds a publicstation by a given id

200

Response body


```
"id": 6,
"electricityPrice": 0.292,
"provider": {
  "address": "na",
  "name": "Stadt Karlsruhe",
  "phone": "na",
  "website": "na"
},
"position": {
  "latitude": 49.005493,
  "longitude": 8.419481
},
"available": false,
```


Figure 6: GET Request for a Specific Charging Station

Figure 7 illustrates the map view as the central frontend element of ECC. Charging stations that are in close proximity are clustered, as shown by the numbered black dots on the map.

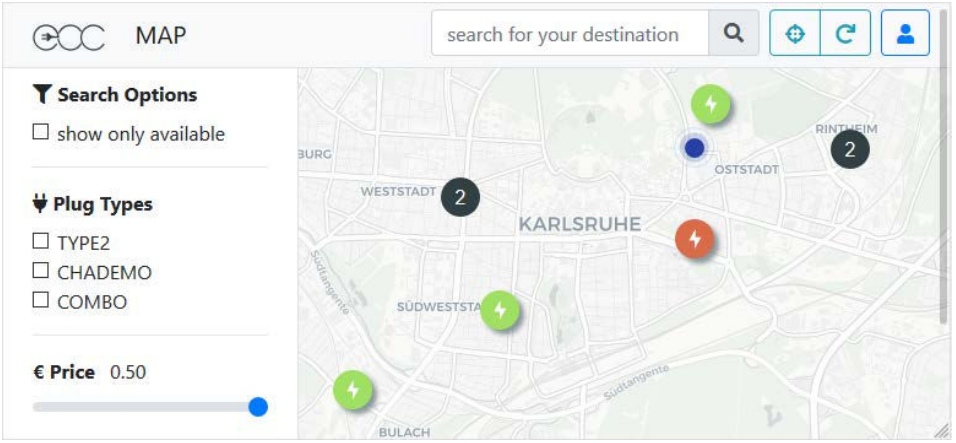


Figure 7: Map View of the ECC Application

Depending on the zoom level, the points are clustered together. When the zoom level is increased the clusters are resolved. Single charging stations are indicated by the bolt icon. An orange bolt icon means that the charging station is currently not available, whereas a green icon indicates a currently available charging station. The blue dot represents the current position of the user. The filter options described in Section 6.1 can be applied in the left menu. Further and detailed information about a charging station is available when one clicks the icon of a charging station.

6.3 Benefits of the Context Map for the Development of Microservices

A context map introduces a structure of a domain that is elaborated by domain experts over a long period of time. The knowledge of the domain is processed in a way that the microservice architecture can be derived from the context map and the microservices from the including bounded contexts. Whenever a new application adhering to the domain should be developed the software development team benefits from the domain knowledge captured by the context map. In the example of the Electric Car Charger, the connected car context map provides a primary architectural structuring of this application (e.g. VehicleManagement including Vehicle, DrivingManagement including ChargingStation, etc., see Figure 3) in a way that the connected car application, ECC, fits into the overall domain structure, and thus into the overall connected car microservice architecture derived from this structure. This enables the re-use of microservices that were implemented during the development of former connected car applications. In our specific case, before the ECC application, a car sharing application was developed which, among others, required the implementation of the Vehicle and the Driver bounded context as microservices. Since the car sharing application and the ECC application are based on the same connected car context map, they can share parts of the map, in this specific case microservices related to the vehicle and the driver. The more applications based on the context map are developed the more microservices can be reused by newly developed applications.

7 Conclusion

A sound understanding of the domain for which a software application should be developed is necessary. A misunderstanding of the stakeholders who should have the domain knowledge and the developer of the software is the main reason why software projects often fail [Sm15]. In the connected car domain a common understanding is constantly growing because there is a high demand on flexible and environmentally friendly mobility solutions. So far, this understanding is documented in an informal way mostly in white papers from companies. We presented an approach on how to formalize the domain knowledge that is available in the field of connected car. Our approach is based on the widely accepted software design concept of Domain-Driven Design. Since this concept provides no formalization on the level of the modeling language, we extended the (also well accepted) Unified Modeling Language to be able to specify the strategic and tactical modeling parts of the domain model by different diagrams. A central diagram which expresses the main structure of the domain is the context map. In this paper, we proposed an initial draft of a context map for the connected car domain. Certainly, the concrete subdomains and including bounded contexts are subject for further discussions. The real value of our contribution is the systematic and formally sound approach on which the discussion of the domain knowledge with experts from the domain can be started – and documented in a way that this knowledge can be directly used in a structured development process. We believe that the close connection of domain knowledge

capturing (also called knowledge crunching) with the software development process is a main advantage of our approach.

We demonstrated our approach with the example of the microservice-based software system Electric Car Charger. We have shown how the context map becomes a central design artifact of the software development process. The context map expresses the main structure of the domain and makes sure that the independently developed microservices are fitting into an overall connected car service landscape. Our approach guarantees that the model and its implementation are always in sync – according to our practical experience this is one of the most important demands of Domain-Driven Design. So far, the alignment of model and implementation is mainly done manually leaving room for model-to-code and code-to-model automation.

References

[Br03] Manfred Broy: Automotive Software Engineering. 25th International Conference on Software Engineering, 2003.

[BG+09] Manfred Broy, Mario Gleirscher, Stefano Merenda, Doris Wild, Peter Kluge, Wolfgang Krenzer: Automotive Architecture Framework: Towards a Holistic and Standardised Sys-tem Architecture Description, Technical Report of the of the Technische Universität München and White Paper of the IBM Cooperation, June 2009.

[Co+16] Riccardo Coppola, Maurizio Morisio: Connected Car: Technologies, Issues, Future Trend. ACM Computing Surveys, Vol. 49, No. 3, Article 46, Publication date: October 2016.

[DK12] Vivek Diwanji; Nilesh Karamarkar: Exploring the Connected Car, Whitepaper, cognizant. 2012, URL: https://www.cognizant.com/InsightsWhitepapers/Exploring-the-Connected-Car.pdf, [retrieved: 2019.04.02].

[DK+18] Soumya Kanti Datta, Mohammad Irfan Khan, Lara Codeca, B. Denis, Jerome Haerri, Christian Bonnet: IoT and Microservices Based Testbed for Connected Car Services, IEEE 19th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM), p. 14 – 19, 2018.

[Ev03] Eric Evans, Domain-Driven Design: Tackling Complexity in the Heart of Software. Addison-Wesley Professional, 2003.

[Fi00] Roy T. Fielding.: Architectural Styles and the Design of Network-based Software Architectures, University of California, Irvine, Dissertation, 2000. https://www.ics.uci.edu/~fielding/pubs/dissertation/fielding_dissertation_2up.pdf, [retrieved: 2019.04.02].

[HG+17] Benjamin Hippchen, Pascal Giessler, Roland H. Steinegger, Michael Schneider, Sebastian Abeck: Designing Microservice-Based Applications by Using a Domain-Driven Design Approach, International Journal of Advances in Software, ISSN 1942-2628, vol. 10, no. 3&4, pages 432 - 445, 2017

[HS+19] Benjamin Hippchen, Michael Schneider, Iris Landerer, Pascal Giessler, Sebastian

Abeck: Methodology for Splitting Business Capabilities into a Microservice Architecture: Design and Maintenance Using a Domain-Driven Approach, Conference on Advances and Trends in Software Engineering (SOFTENG 2019), Valencia, 2019.

[KA+16] Per-Henrik Karlsson, Hong K. Ahn; Byeongmin Choi: Connected Car – A New Ecosystem, Ipsos Business Consulting. 2016, URL: https://www.ipsos.com/sites/default/files/2016-06/022.1-connected-car-a-new-ecosystem.pdf, [retrieved: 2019.04.02].

[MK+17] Arthur de M. Del Esposte, Fabio Kon, Fabio M. Costa, Nelson Lago: InterSCity- A Scalable Microservice-based Open Source Platform for SmartCities, In Proceedings of the 6th International Conference on Smart Cities and Green ICT Systems (SMARTGREENS), pages 35-46, 2017.

[Ne15] Sam Newman: Building Microservices, O'Reilly Media, Inc., 2015.

[PK+16] Patrizio Pelliccione, Eric Knauss, Rogardt Heldal, Magnus Agren, Piergiuseppe Mallozzi Anders Alminger, Daniel Borgentun: A proposal for an Automotive Architecture Frame-work for Volvo Cars, IEEE Workshop on Automotive Systems/Software Architectures, 2016.

[SH+18] Michael Schneider, Benjamin Hippchen, Sebastian Abeck, Michael Jacoby, Reinhard Herzog: Enabling IoT Platform Interoperability Using a Systematic Development Approach by Example, Global Internet of Things Summit (GIoTS). IEEE, 2018. pages 1 – 6, 2018.

[Sm15] John Ferguson Smart: BDD in Action – Behavior-Driven Development for the whole software lifecycle. Manning Publications, 2015.

[TH+18] Shmuel Tyszberowicz, Robert Heinrich, Bo Liu, and Zhiming Liu: Identifying Microservices Using Functional Decomposition. International Symposium on Dependable Software Engineering: Theories, Tools, and Applications. Springer, Cham, 2018.

[VA+14] Richard Viereckl, Jörg Assmann; Christian Radüge: In the fast lane – The bright future of connected cars, strategy&., 2014, URL: https://de.scribd.com/document/379805993/Strategyand-In-the-Fast-Lane-pdf, [retrieved: 2019-04-02].

[Ve13] Vaughn Vernon: Implementing Domain-Driven Design. 1st. Addison-Wesley Professional, 2013.

Combining the Concepts of Semantic Data Integration and Edge Computing

Matthias Farnbauer-Schmidt[1,2], Julian Lindner[3], Christopher Kaffenberger[4], Jens Albrecht[5]

Abstract: The *Internet of Things* (IoT) is growing rapidly. Therefore, there are more and more vendors, which led to IoT being a heterogeneous collection of different IoT platforms, isolated solutions and several protocols. It has been proposed to use Data Integration to overcome this heterogeneity. In addition, costs are on the raise due to increasing volume of data which increases demands on bandwidth and cloud computing capabilities. Again a solution has already been proposed by reducing the amount of data to forward by processing data at the edge of an IoT-System, e. g. filtering or aggregation. This concept is called Edge Computing.

In this article the Semantic Edge Computing Runtime (SECR) is introduced, combining both concepts. The application of Data Integration enables Edge Computing to be performed on a higher level of abstraction. In addition, the developed Driver-approach allows SECR's Data Integration algorithm to be applied to a wide range of data sources without imposing requirements on them. The Data Integration itself is based on technologies of Semantic Web, applying metadata to raw data giving it context for interpretation. Furthermore, SECR's REST-API enables applications to alternate Data Integration and Edge Computing at runtime.

The tests of SECR's prototype implementation have shown its suitability for deployment on an edge device and its scalability, being able to handle 128 data sources and Edge Computing Tasks.

Keywords: Internet of Things; Data Integration; Edge Computing; Semantic Web; SECR

1 Introduction

The *Internet of Things* (IoT) is the approach of linking the real world to the Internet. Consequently, digitalization of real-world properties is done by measurements conducted by sensors.

The dominant architecture of IoT applications relies on a central cloud, i. e. a powerful computational center. All data produced by sensors is forwarded to the cloud for processing,

[1] Technische Hochschule Nürnberg Georg-Simon-Ohm, Keßlerplatz 12, 90489 Nürnberg, Germany
[2] Fraunhofer IIS Arbeitsgruppe SCS, Nordostpark 93, 90411 Nürnberg, Germany farnbams@scs.fraunhofer.de
[3] Fraunhofer IIS Arbeitsgruppe SCS, Nordostpark 93, 90411 Nürnberg, Germany julian.lindner@scs.fraunhofer.de
[4] Fraunhofer IIS Arbeitsgruppe SCS, Nordostpark 93, 90411 Nürnberg, Germany christopher.kaffenberger@scs.fraunhofer.de
[5] Technische Hochschule Nürnberg Georg-Simon-Ohm, Informatik, Keßlerplatz 12, 90489 Nürnberg, Germany jens.albrecht@th-nuernberg.de

doi:10.18420/inf2019_19

storing and decision making. Furthermore, there are Gateways that translate protocols on the way from data source to cloud.

Scalability problems of the cloud-centric IoT-architecture are pointed out by the growing number of devices [Ga17]. As a result, the more devices are deployed the more data is produced. With an increased volume of data, a cloud requires higher computational resources. Moreover, the network connecting devices and cloud must provide a higher bandwidth to be able to convey it. In fact, bandwidth is a constraint resource and both, computational power and bandwidth are expensive. A solution to this problem is introduced by Edge Computing where data is pre-processed at the edge.

The IoT is highly heterogeneous today [Qi18]. It can be seen at every layer of the ISO-OSI-model. In addition, the representation of data within a protocol can be heterogeneous either. For instance, there can be differences in units, scale and meaning. In fact, temperature of 32 could mean 32 m°Cor 32 K and could be the room temperature or the average temperature in space. Representation is usually defined by contract at protocol, platform or application level. Besides, some domains have their own niche solutions. A proper way to overcome heterogeneity of different data sources is to perform Data Integration. The approach of using Semantic Web Technology has been introduced to the IoT and is called Semantic Web of Things.

Edge Computing requires Data Integration when computations should be applied to data from different sources. In order to produce sensible results, computations require their inputs to be modeled according to the same schema. In fact, the Data Integration decouples the execution of computations from the heterogeneity of data sources. As a result, Edge Computing software that includes a Data Integration layer is more reusable than Edge Computing software that handles specific data sources.

The benefits of combining Semantic Data Integration and Edge Computing will be shown by introducing the Semantic Edge Computing Runtime (SECR). Focused on performing pre-processing for data science algorithms, it works as a backend for IoT-applications on the edge. SECR is designed to be deployed at edge devices at least capable of running an OS. This excludes the outermost edge devices like simple sensors and actuators. An abstraction of data sources in combination with the developed Driver-approach allows SECR's Semantic Data Integration to handle a wide range of data sources. In addition, a local RDF-graph is maintained that provides all information of SECR, its host and environment. It is internally used for configuration of services, either. Furthermore, a REST-API is provided to access the graph. Moreover, the API allows for modification of Edge Computing tasks and Semantic Data Integration at runtime.

2 Background

This section addresses the solutions to the problems of IoT before mentioned. In addition, their background and technologies are covered.

2.1 Data Integration

Heterogeneity of data sources can be overcome by applying Data Integration. It is done by transforming data into a common schema. As a consequence, all integrated data can be queried as a whole. A schema is a description of how certain information is modeled. Although, a schema only defines the semantics of a data model not the syntax the data is represented in.

Data Integration enables interoperability if the communicators understand the common schema. The lower the system-layer Data Integration is applied the earlier interoperability between IoT-systems can be achieved.

2.2 Semantic Web

The Semantic Web or Web of Data wants to interlink the data provided in the Internet. This concept is called Linked Data [LPL17].

The standard used for Linked Data is the Resource Description Framework[6] (RDF) a recommendation of the World Wide Web Consortium (W3C). The Framework sees the description of information in subject-predicate-object-triples, e. g. “Hans is male”. Subjects and predicates must be resources identified by an Uniform Resource Identifier (URI) whereas objects can be either a resource or a literal. Several interlinked RDF-triples build a directed graph where subjects and objects are the nodes and the predicates are the directed edges.

2.2.1 Ontologies

An ontology describes entities and the relations between them. In case of the Semantic Web an ontology is defined by RDF-statements (RDF-triples). These statements are divided into two groups the terminological box (TBox) and the assertion box (ABox) [Bo17]. The TBox-statements provide classes and predicates to identify entities and their kind of relations. In contrast, the ABox-statements use the terms defined by the TBox to describe entities and their relations.

Depending on the share of TBox- and ABox-statements, ontologies are either classified as vocabulary or as knowledge-graph in this paper. This is done in order to express the purpose of an RDF-graph.

The TBox-statements of a vocabulary define a schema for modeling data. They can describe a broad domain or extend such a vocabulary into more detail. An example is the Semantic Sensor Network (SSN) ontology[7] that exapnds the Sensor, Observation, Sample and Actuator

[6] https://www.w3.org/RDF/
[7] https://www.w3.org/TR/vocab-ssn/

(SOSA) ontology. Both are ontologies provided by the W3C. The term ontology is often used as a synonym for vocabulary.

Knowledge-graphs use vocabularies to model entities and their relations. By using a common vocabulary the semantics of a graph can be understood by everyone that knows the vocabulary.

2.2.2 Data Integration by Application of Vocabularies

Building knowledge graphs by using the terms of vocabularies is a kind of Data Integration, in the future referred to as Semantic Data Integration (SDI). The schema built from a vocabulary's statements works as a common schema for Data Integration. Being a directed graph, the linked statements of an RDF-ontology can be traversed. Therefore, if a reader knows the vocabulary used to describe the entities of a graph he is able to infer the semantics of that graph.

2.3 Edge Computing

Edge Computing tackles IoT's issue of an increasing volume of data. It utilizes the execution of computations on edge devices. The concept leverages the computational powers of devices of the outer ends of an IoT-system to reduce the payloads for network and cloud.

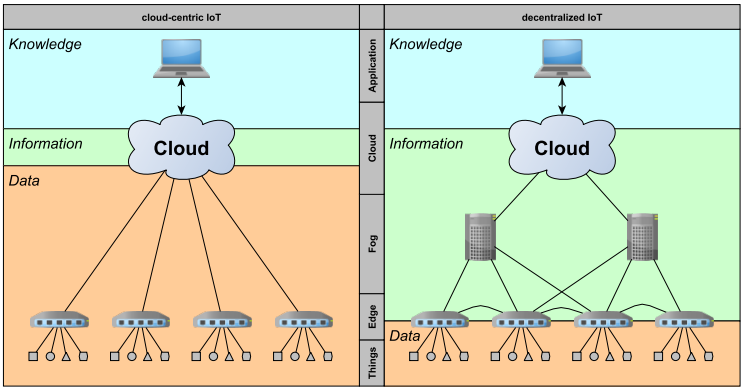


Fig. 1: Cloud-centric IoT-architecture (left) and decentralized IoT-architecture including Edge Computing (right). Conversion from Data to Information takes place at gateway level. Redrawn according to [Pa17].

The term Edge Computing is not globally defined. The definition of the "edge" used in this article is shown in Figure 1. It is composed of the gateways connecting sensors and actuators to the infrastructure of the IoT. Therefore, the Edge Computing introduced by

SECR only targets those edge devices. For example, Brown, Kathrivel and Akthar define Edge Computing to be executed on decentral micro-clouds [Br17; KA17] (see Fog in Figure 1). Whereas others see the edge as the outermost sensors and actuators (see Things in Figure 1).

The distribution of computations increases the complexity of an IoT-system because the cloud is no longer the only instance that conducts computations. Edge devices are heterogeneous and provide different levels of computational powers. As a consequence, a system has to be introduced to manage deployment of Edge Computing and load balancing. Such systems are referred to as Edge Computing platforms. In contrast, the software deployed at an edge device to execute Edge Computing is called Edge Computing software. An Edge Computing platform distributes tasks requested by an application to edge devices deploying Edge Computing software. After all, depending on the complexity of a system and the number of edge devices Edge Computing platforms are optional.

3 Related Work

Semantic Data Integration and data pre-processing has been suggested by other projects before and will be examined in the following.

Desai et al. introduced the concept of Semantic Gateway as a Service [Al15]. The purpose of it is to break up *vertical silos*. These are closed IoT-applications which are obstacles on the way to enable gateway-level interoperability. In fact, the concept is limited to Semantic Data Integration. Applications can access the results either by push-pull REST-API or by event-driven MQTT. Besides, a multi-protocol proxy is used for handling of data sources. In a Semantic Gateway the data sources must implement a certain protocol. So, the sources itself provide descriptions of their packages for the gateway. These descriptions are used to extract the contained data of a package. All in all, the requirement for data sources to implement a protocol limits the data sources that can be handled by the Semantic Gateway.

Semantic enrichment of data causes an increase of payload due to additional metadata. Al-Osta et al. further developed the concept of Semantic Gateway to reduce the data that must be forwarded by pre-processing incoming data [AAA17]. According to this report's definition this is Edge Computing. However, data sources are still required to implement a certain protocol to work within this system. Their Data Preparation Module reduces traffic by applying rules of aggregation and filtering. Consequently, Edge Computing capabilities are restricted. In contrast, SECR provides richer Edge Computing capabilities, allowing for dependencies between sources, scaling and converting of data.

To sum up, Semantic Gateway and its derivatives show that Semantic Data Integration can be done at gateway-level. In addition, Semantic Data Integration and rule-based data processing can not only reduce the emitted information but also create new information.

4 The Semantic Edge Computing Runtime

Designed as a backend for IoT-applications on the edge, SECR is Edge Computing software. Besides, the small footprint leaves enough resources to run further services on the same host. All results of SECR's services are published as RDF-graphs enabling edge-level interoperability. Moreover, SECR's Edge Computing capabilities focus on data processing, e. g. filtering, aggregation, fuzzyfication and classification.

Publish-subscribe HTTP and event-driven MQTT is used for SECR's public API (see Figure 2). The HTTP-API is used to pull results from SECR's services, configure the services and for querying SECR's local RDF-graph whereas MQTT is used for event-driven publication of service results. In MQTT content is published to so-called topics. The protocol is handled by a broker, which notifies all subscribers of a topic when new content is published. The Data Source Managers (DSMs) in Figure 2 are a proxy for the instances that handle the Semantic Data Integration. Similarly, the Edge Computing Tasks (ECTs) do the Edge Computing.

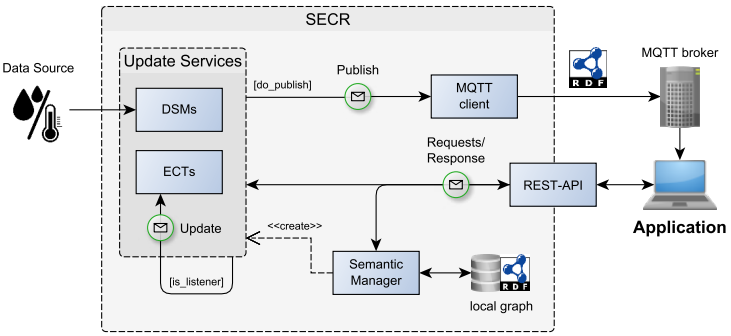


Fig. 2: Architecture of SECR. Applications can either interact with the REST-API or the MQTT-broker. Update Services are launched by the Semantic Manager which holds access to SECR's local RDF-graph. Edge Computing relies on the results of other Update Services.

The components of SECR will be further discussed and explained in the rest of this section.

4.1 SECR's Public Services

Metadata of SECR can be obtained from the public accessible local RDF-graph and from the Resource-Usage-Information service. Provided are memory allocation and CPU load of SECR's host system, the average latencies of Edge Computing and Semantic Data Integration as well as a description of the host system, its environment and SECR's deployed services. For reasons of clarity these services are not shown in Figure 2.

A vital part of SECR are Update Services. Each Update Service handles a Semantic Struct which is SECR's representation of results from Semantic Data Integration or Edge

Computing. An Update Service's purpose is to provide updates of a Semantic Struct to consumers of the service's results. Furthermore, Semantic Structs can be serialized into an RDF-graph. A user can subscribe to a Semantic Struct's state at an MQTT-topic. However, an Update Service must be set to publish to MQTT.

4.2 Semantic Structs

Semantic Structs consist of a timestamp of their last update, the URI of their Update Service and at least one Field identified by a label, an URI and the type of data they hold.

```
@prefix rdf  <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix sosa <http://www.w3.org/ns/sosa/> .
@prefix sec  <http://iis.fraunhofer.de/vocab/sec/> .

_:obs a sosa:Observation ;
  sosa:usedProcedure {Update-Service-URI} ;
  # for each of Semantic Struct's Fields
  sosa:hasResult [
        sec:instanceOF {Field-URI} ;
        rdf:value {current-value-of-Field}
  ] ;
  sosa:resultTime {Semantic-Structs-timestamp as
       YYYY-MM-DDThh:mm:ss.sss} .
```

Fig. 3: Pattern of an RDF-graph representing the state of a Semantic Struct. The Turtle-syntax is used.

The state of a Semantic Struct is returned as an RDF-graph to users. The RDF-graph resembles a `sosa:Observation` following the pattern of Figure 3. The chosen graph-pattern provides all information necessary to discover the full semantic description of the Semantic Struct and the changing values of Fields and time of update.

4.3 Semantic Data Integration

Semantic Data Integration is performed by a combination of Drivers, Data Source Managers (DSMs) and Semantic Conversion Services (SCSs). The components of SECR's Semantic Data Integration layer are shown in Figure 4 and described in the following paragraphs.

Data Sources are an abstraction used by SECR's Semantic Data Integration to handle IoT's heterogeneous data sources. A *Data Source* is able to emit packages of raw data. These packages are called Frames which consist of different Fields similar to Fields of a Semantic Struct. The Fields of a Frame hold the raw data.

For each *Data Source* SECR handles, a full semantic description is provided in SECR's local RDF-graph. Different *Data Sources* can be of the same type, e. g. several sensors (entities)

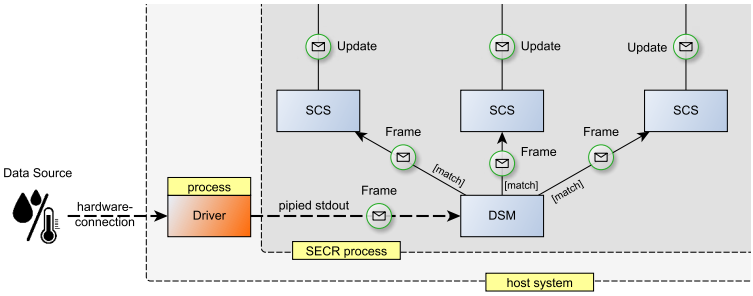


Fig. 4: SECR's Semantic Data Integration of a *Data Source* is a collaboration of several components.

of the same model (type) could be deployed. A type's description defines the Driver to be used and which Frames are emitted and their properties. Furthermore, a Frame's description includes its URI, a label, a pattern that uniquely identifies the Frame, the description of its Fields and information how to extract the data of the Fields. Accepted Frame formats are byte-arrays, ASCII-strings, XML and Json. A Frames pattern depends on its format. For example, the pattern of a byte-array Frame is a sequence of bytes that must match the Frame's content. In contrast, the pattern of a Json-format Frame is a key-value-pair that must be contained in the Json. As of Semantic Structs, Fields of a Frame are defined by an URI, a label and their type of data. The encoding of the Fields as well as the pattern depends on a Frame's format. For example for byte-array Frames a Field is localized by position and length in the array whereas Json- and XML-encodings provide the keys where the data is found.

The developed **Driver**-approach allows SECRs to abstract over IoT's heterogeneous data sources. A Driver's purpose is to validate a *Data Source*'s protocol, to handle the connection between the *Data Source* and the SECR and to forward validated Frames to its DSM. For each *Data Source* handled by SECR one Driver is launched. As depicted in Figure 4 a Driver is a child-process launched by SECR. Therefore, Drivers can be configured by passing command-line-arguments at their launch. Arguments can be specified for each *Data Source type*, for each *Data Source* and at setup of a DSM.

As an example, we assume a HTTP-server as *Data Source*. Therefore, the Driver would be responsible to continuously request new Frames from the server. In addition, the response bodies contain a custom checksum that must be validated by the Driver to proof a Frames validity.

A **Data Source Manager** (DSM) coordinates the conversion of Frames into Semantic Structs. This is done by receiving validated Frames from the Driver and passing them to the responsible Semantic Conversion Service. Therefore, it is the DSM's task to determine which kind of Frame is received by applying the patterns of the handled *Data Source*'s Frames.

For each Frame a *Data Source* can emit a **Semantic Conversion Service** is deployed by the Data Source Manager. They are Update Services whose task is to convert a Frame into their Semantic Struct. The conversion is done by extracting the raw data of the Frame and updating the corresponding Fields of the SCS's Semantic Struce. For extraction the encoding information from the Frame's description is used.

4.4 Edge Computing Tasks

Edge Computing Tasks (ECTs) are SECR's source of Edge Computing capabilities. Consequently, each ECT is an Update Service. Indeed, the Edge Computing is done by calculating new values for the Fields of an ECT's Semantic Struct. For each Field an expression is provided. SECR's supported types of data are *Numeric*, *Boolean* and *Categorical*. Furthermore, a Boolean expression called `publish-condition` is provided for each ECT. It defines the moments when an ECT updates the state of its Semantic Struct.

4.4.1 Setup of an ECT

The creation of an ECT requires a SECR-wide unique label, description of the new Semantic Struct's Fields, the expression for them and the `publish-condition`. Accordingly, all required information must be provided except for the ECT's label which is encoded in the HTTP-request's URL.

For example we want to create a new ECT called `task` on a SECR (`${base} = http://example.org:8080/secr0`). On the SECR a SCS (`${base}/dsm/other/scs/frame`) and an ECT (`${base}/ect/another`) are already running. Field `x` of the new ECT should be calculated from the value of SCS's Field `a` and 30 and Field `y` should be calculated from the value of `x` and value `t` of the other ECT. At last, updates should be published when `x` exceeds 50. We achieve the described behaviour by sending the Json-LD from Figure 5 to `POST http://example.org:8080/secr0/ect/task`.

Unlike most of the required semantics, the expressions are not encoded as RDF (see Figure 5). This decision was made for user friendliness because complex expressions would result in large and complex RDF-graphs.

Fields in expressions are accessed by `{identifier}#{FieldLabel}` where an identifier is either the URI of another Update Service or `SELF` which indicates that the Field is part of the ECT's own Semantic Struct. Own Fields can only be referenced when they have been declared before, e. g. Field `x` could not reference Field `y`.

```
// replace ${base} by http://example.org:8080/secr0

{ "@context":"${base}/context.json",
  "dependencies": [
        { "other":"${base}/dsm/other/scs/frame" },
        { "another":"${base}/ect/another" }
  ],
  "fields": [ {
          "label":"x", "ofType":"Numeric",
          "expression":"<other#a> + 30"
        }, {
          "label":"y", "ofType":"Numeric",
          "expression":"<another#t> + SELF#x"
        } ],
  "publish_when": { "expression":"<SELF#x> > 50" }
}
```

Fig. 5: Example content of POST to create a new Edge Computing Task. The Semantic Struct will consist of Fields `x` and `y`. The ECT will depend on SCS `other/frame` and ECT `another`.

4.4.2 Algorithm of Evaluation

The evaluation of ECT's expressions is driven by the updates of the services they depend on. For each dependency an execution plan is created, e. g. the execution plan for the example ECT `task` is shown in Table 1. Indeed, the Update Service `other` is only mentioned in the expression for Field `x` (see Figure 5). However, the execution plan for updates from `other` additionally recalculates Field `y` and the `publish-condition` because they depend on Field `x`.

Tab. 1: Resulting execution plans from the instruction of Figure 5.

Update from	**Execution plan**
other	Recalculate `x` → recalculate `y` → recalculate `publish-condition`
another	Recalculate `y`

5 Evaluation

The prototype implementation of SECR is tested for scalability and suitability for deployment at the edge.

For evaluation purposes, a special Driver has been implemented. The Driver itself simulates a *Data Source* that emits every 50 ms one Frame. Furthermore, the *Data Source* can send three different kinds of Frames which one is sent is determined by chance.

5.1 The Test Scenario

For the tests SECR is deployed on a RaspberryPi 3 Model B that runs a quad-core Arm-processor at 1.2 GHz.

For testing the scalability several tests are run with 2 up to 128 (2, 4, 8, 16, 32, 64, 96, 128) simulated *Data Sources* at one time. All tests are run for 120 s 30-times. Four test cases have been evaluated:

1. n DSMs are deployed; Nothing is published to MQTT.
2. n DSMs are deployed; All results are published to MQTT.
3. n DSMs and $n - 4$ ECTs are deployed; Nothing is published to MQTT.
4. n DSMs and $n - 4$ ECTs are deployed; All results are published to MQTT.

In the third and fourth case the ECTs depending on four SCSs, created from instructions like the example in Figure 6. Indeed, the fourth test case can be seen as a worst-case scenario.

```
// replace ${base} by http://localhost:8080/secr0

{ "@context":"${base}/context.json",
  "dependencies": [
        { "s31_temp":"${base}/dsm/S31/scs/temp" },
        {  "s32_vel":"${base}/dsm/S32/scs/vel" },
        { "s33_temp":"${base}/dsm/S33/scs/temp" },
        { "s34_temp":"${base}/dsm/S34/scs/temp" }
  ],
  "fields": [ {
          "label":"temp_avg", "ofType":"Numeric",
          "expression":"mov_avg(4, s31_temp#t * 0.01)"
        }, {
          "label":"mul", "ofType":"Numeric",
          "expression":"s32_vel#v * s33_temp#t"
        } ],
  "publish_when": { "expression":"s34_temp#t.ROSE" }
}
```

Fig. 6: Body of `POST http://localhost:8080/secr0/ect/alert34`. Instruction to create ECT `alert34` for the evaluation of SECR's prototype implementation.

Time and resource consumption measurements are part of SECR's Resource-Usage-Information service. Therefore, the measurements do not generate any extra costs.

The latencies of the Update Services where measured to determine the payload possible to be handled by SECR running on a RaspberryPi 3. On the one hand, the latency of Semantic Data Integration defines the time elapsed from the moment a new Frame is read from the

Driver's `stdout` to the moment the serialized RDF is sent to the MQTT-broker. On the other hand, the latency of Edge Computing is defined as the time elapsed from the moment the ECT received the notification of update to the moment the serialized RDF is sent to the MQTT-broker.

5.2 Results and Discussion

The results of the tests are shown in Figure 7 through Figure 10. Note the non-linear x-axis. In addition, it should be considered that due to other processes the latencies can be disturbed. Therefore, the results are presented as boxplots.

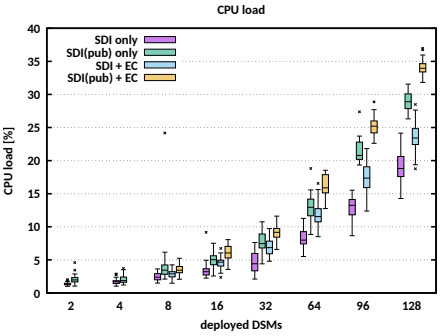


Fig. 7: CPU load of the RaspberryPi 3 when running SECR.

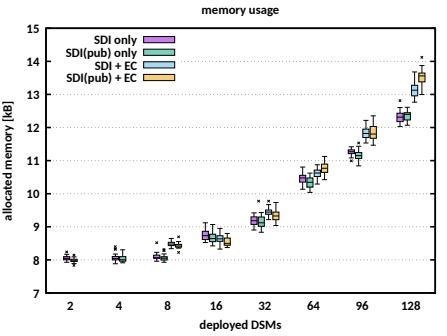


Fig. 8: Allocated memory of SECR's process on the RaspberryPi 3.

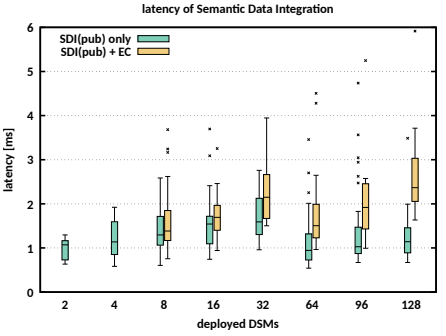


Fig. 9: Latency of Semantic Data Integration.

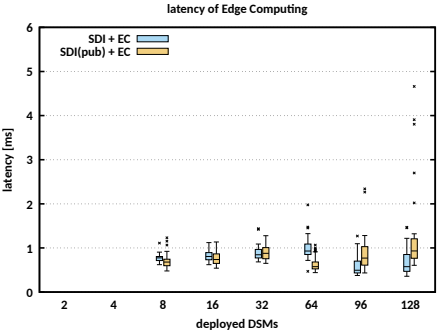


Fig. 10: Latency of Edge Computing.

As expected, in each test case the CPU load rises nearly linear with the number of *Data Sources* handled. By capturing one third of the hosts CPU in a worst-case scenario, SECR leaves enough resources to run further processes achieving the objective defined in section 4.

The memory allocation of SECR's process starts with an offset of 8 kB and rises in all test cases nearly linear. The offset is mainly derived from SECR's local RDF-graph. By

allocating 13.5 kB in test case 4, SECR is suitable to be deployed at systems with limited memory capacities.

The latency of Semantic Data Integration shows a number of outliers but stays generally below 4 ms. The latencies of both Semantic Data Integration and Edge Computing rise with the number of deployed services due to emerging dependencies between them. In the worst-case a new Frame is integrated within 6 ms and processed by an ECT within 5 ms. Finally, the sum of both propagation times is nearly 5-times faster than the occurrence of new Frames.

6 Conclusion and Future Work

In this paper the authors introduced the Semantic Edge Computing Runtime (SECR). The software enables Edge Computing capabilities on the host system. The heterogeneity of the Internet of Things (IoT) is handled by applying Data Integration before the Edge Computing. Besides, SECR's Data Integration is done by converting raw data into RDF-graphs. This abstraction before the computations allows SECR to apply Edge Computing on a wide range of *Data Sources*.

The evaluation of the prototype SECR has proven its suitability for deployment at the edge. Compared to the results of [AAA17] the overall worst latency of 11 ms is 3-times faster than their average latency of 30 ms.

In the future we will examine a real-world use-case to investigate the effects of Edge Computing in terms of saving bandwidth. In order to make SECR production ready concerns are taken towards security and failure safety.

With more computations taking place at the edge the interest of attackers raises. Therefore, measures must be taken to prevent malicious attacks. Nevertheless, edge devices are resource constraint. So, a compromise must be found between resource consumption and safety measures.

The loss of functionality after failure must be prevented. Currently, SECR provides no persistence of its services. Certainly, edge devices are more prone to failures than computational centers. In addition, it is advantageous to turn of edge devices to save energy sometimes. On restart SECR should restore the state of its services.

7 Acknowledgement

This work was partially supported by the Bavarian State Ministry of Economic Affairs, Regional Development and Energy within the framework of the Bavarian Research and Development Program "Information and Communication Technology".

References

[AAA17] Al-Osta, M.; Ahmed, B.; Abdelouahed, G.: A Lightweight Semantic Web-based Approach for Data Annotation on IoT Gateways. International Conference on Emerging Ubiquitous Systems and Pervasive Networks 8th/, 2017.

[Al15] Semantic Gateway as a Service Architecture for IoT Interoperability. In (Altintas, O., ed.): 2015 IEEE International Conference on Mobile Services (MS). IEEE, Piscataway, NJ, pp. 313–319, 2015, ISBN: 978-1-4673-7284-8.

[Bo17] Bonte, P.; Ongenae, F.; Backere, F.; Schaballie, J.; Arndt, D.; Verstichel, S.; Mannens, E.; Walle, R.; Turck, F.: The MASSIF Platform: A Modular and Semantic Platform for the Development of Flexible IoT Services. Knowl. Inf. Syst. 51/1, pp. 89–126, 2017, ISSN: 0219-1377, URL: https://doi.org/10.1007/s10115-016-0969-1.

[Br17] Brown, K.: Resiliency of Edge Data Centers in the Era of Cloud Computing, YouTube, 2017, URL: https://www.youtube.com/watch?v=ttto-t4asE0, visited on: 07/31/2018.

[Ga17] Gartner, I.: Gartner Says 8.4 Billion Connected Things Will Be in Use in 2017, Up 31 Percent From 2016, 2017, URL: https://www.gartner.com/en/newsroom/press-releases/2017-02-07-gartner-says-8-billion-connected-things-will-be-in-use-in-2017-up-31-percent-from-2016, visited on: 11/07/2018.

[KA17] Kathirvel, K.; Akhtar, H.: Implications of 5G and Edge Computing on OpenStack, Youtube, 2017, URL: https://www.youtube.com/watch?v=9d5JtONGQSA, visited on: 07/31/2018.

[Ka18] Kaed, C. E.; Khan, I.; van den Berg, A.; Hossayni, H.; Saint-Marcel, C.: SRE: Semantic Rules Engine for the Industrial Internet-Of-Things Gateways. IEEE Transactions on Industrial Informatics 14/2, pp. 715–724, 2018, ISSN: 1551-3203.

[LPL17] Li, W.; Privat, G.; Le Gall, F.: Towards a Semantics Extractor for Interoperability of IoT Platforms. Global Internet of Things Sumit/, 2017.

[Pa17] Pande, A.: IOT Edge Computing | IoT Examples | Use Cases | HackerEarth Webinar, YouTube, 2017, URL: https://www.youtube.com/watch?v=Xm8frqTZRVI, visited on: 07/31/2018.

[Qi18] Qiu, T.; Chen, N.; Li, K.; Atiquzzaman, M.; Zhao, W.: How Can Heterogeneous Internet of Things Build our Future: A Survey. IEEE Communications Surveys & Tutorials/, p. 1, 2018.

Derivation of Categories for Interoperability of Blockchain- and Distributed Ledger Systems

Katharina Zeuch, Kai Hendrik Wöhnert, Volker Skwarek[1]

Abstract: Due to increasing security requirements e. g. for transaction based smart-x-technologies in distributed systems, blockchain technologies are predestined for secure data exchange and keeping in distributed systems. Although the underlying principle of almost every blockchain is the Byzantine fault tolerance (BFT), its implementation differs significantly between the technologies so that migration or interoperability between systems is nearly impossible. Additionally, this missing interoperability also reduces the chance for scalability between different extents of implementation as there is usually not a one-size-fits-all-blockchain: Different technologies have their advantages for different systems. Therefore scalability and interoperability are tightly coupled. As a basis for further research on and the derivation of generally scalable and interoperable architectures of blockchains, current technologies have to be made comparable and interoperability criteria have to be developed. This paper analyses current literature and introduces technical criteria for the comparison of blockchain- and distributed ledger technologies (BC/DLT).With a list of eleven criteria popular BC/DLTs such as Bitcoin, Ethereum, Hyperledger Fabric, Ripple and Corda are compared regarding general features.

Keywords: blockchain and distributed-ledger technology (BC/DLT); comparison categories; abstraction; scalability; interoperability

1 Introduction

For data integrity and security in distributed systems, blockchain and distributed ledger technologies (BC/DLT) have developed to the state-of-the-art during recent years. Various BC/DLT have been developed and introduced to the public such as Bitcoin, Ethereum or Hyperledger. They provide increased security to and trust into the integrity and authenticity of data in distributed systems by cryptographic linking and consensus mechanisms.

Examples for blockchain applications can be found in many sectors: Benefits are used, e.g. for pharmaceutical tracking and tracing [Ku18; Me16], for healthcare management [DCP18]. Further BC/DLT are proposed to be used in the social sector e. g. the implementation of digital elections as described by [KV18] or [CWB18]. Also for industrial applications considering the Internet-of-Things (IoT) BC/DLT offers new business opportunities. [Hu16] or [FF18] describe scenarios of future IoT, using blockchain for daily applications. Examples for automotive applications of blockchain technology are blockchain-based car insurance systems [La18] or blockchain-based selection of charging stations for electric cars [PKS16].

[1] Hamburg University of Applied Sciences, Forschungs- und Transferzentrum *Digitale Wirtschaftsprozesse*, Forschergruppe *DLT*$^{3}{}_{Hamburg}$, Faculty Life Sciences, Ulmenliet 20, 21033 Hamburg, Germany, firstname.surname@haw-hamburg.de

doi:10.18420/inf2019_20

Due to the variety of available blockchain solutions, an initial decision for one specific blockchain for a specific application appears to be difficult. Missing concepts for scalability and interoperability prevent data from being transferred or migrated between blockchains. Although some concepts for interoperability have meanwhile been developed with technologies such as Interledger [TS], Overledger [Ve18], Polkadot [Wo] or Cosmos [Te] their acceptance in terms of a distributed system remains critical. They are usually introducing central instances, managing transactions between the blockchain- and distributed ledger systems, bypassing its decentrality. Therefore a universal approach based on *interoperability-by-architecture* in terms of a generic description of blockchain characteristics and properties is needed. Such generic systems, which are still subject to research, can behave like different blockchains basing on its configuration.

For the first step to such an architecture, specific properties of blockchain systems have to be discovered, examined and compared. Such criteria will be derived in this article by literature research. For example, [TT17] created a comprehensive taxonomy on blockchain technology and various comparisons of different blockchains have been executed, e.g. [Hi17; Zh17].

This paper will derive general categories for the comparison of blockchain and distributed ledger technologies. In section 2 technical core principles of BC/DLT and its interoperability will be introduced, before the categories will be developed in section 3.

2 Blockchain-Technology

About ten years after the publication of Bitcoin by Nakamoto [Na09], a wide variety of BC/DLT and related principles can be observed in practical use. Although the underlying technology bases on BFT-mechanisms, practical implementations differ significantly between the different technologies. As there is not yet a generally agreed definition of the terms *blockchain or DLT*, this article refers to the following core principles as essential characteristics for this technology:

block building and chaining by cryptographic lining Unlike general DLT-systems, blockchains are characterized by aggregating verified and valid transactions into a block by miners [Bu19]. As these blocks may also contain false information injected by malicious miners, the miner's role is more sensitive for the complete system security. Although generated blocks are later verified and potentially rejected by other network-participants again, this is a known attack to blockchains - e.g. by selfish mining [Li17].

According to the Byzantine generals problem, this content proposing role is the only which can create information whereas others can only verify. Therefore, special security mechanisms are installed for such sensitive instances, so they cannot be abused by hijacking or infiltrating the with malicious code. In some systems the information creation itself is distributed to multiple peers such as in Hyperledger

Fabric, in other systems the role is randomly switched among qualified (mining) nodes such as in Bitcoin or Ethereum.

Part of this block is the reference to the preceding block, usually represented by the previous block-hash-value. This information forms the chain of blocks by cryptographic links and makes it immutable.

distribution Every peer in a distributed blockchain network potentially - depending on its role - owns a copy of the blockchain [En19]. Therefore all transactions can be seen and verified by every instance of the blockchain. Changing the blockchain thereby means to change all copies at all nodes.

implicit or explicit consensus Generated blocks have to be verified for correctness by using a consensus process. The community-used term usually mixes different aspects of this consensus process:

- It partly includes a *leadership election* for the more influential block-proposer role such as in the *proof-of-work* algorithm,
- it partly requires explicit consent communication such as performed by the endorsement nodes in Hyperledger Fabric proof-of-workör
- the consensus process is executed independently on all nodes by the same algorithm without explicit communication about the consent result such as in Bitcoins or Ethereums proof-of-work and should therefore be rather called *implicit consensus*.

Building up on these principles diverse BC/DLT are created. However, no application could be called *THE blockchain*. By [va] already over forty projects, that support smart contracts, have been listed. Even more could be added if taking those blockchains into account, that do not natively support smart contracts, like for example Bitcoin.

Apart from the capability of executing smart contracts, further differences occur for example in the field of distribution, where the amount of data stored by network peers can differ. Alternatively, in the area of consensus, where many different mechanisms can be used to create a network consensus. Thus each blockchain is different from the others in special features of the implementation or the set of rules and the proposed field of application and use cases.

Differences between BC/DLTs can also occur in terms of interoperability, the „ability of systems to provide services to and accept services from other systems and to use the services so exchanged to enable them to operate effectively together“ [In17], and scalability. [HLP18] state that interoperability between blockchain-systems will be a core requirement if blockchain technology becomes a fundamental future data infrastructure. The increasing acceptance and usage of BC/DLTs make scalability important. But scalability of blockchain-systems is limited by the blockchain size, the transaction processing rate and data transmission latency [Xu17].

3 Derivation of Comparison Criteria

A multitude of BC/DLT variations exists, that are scalable and interoperable in different dimensions. A decision for one BC/DLT-system should base on a comparison of various BC/DLTs. Such a comparison should take aspects of scalability and interoperability into account, because of their importance. Most BC/DLTs are created as standalone systems. Interoperability between BC/DLT-systems will prevent users from being locked to one chosen blockchain [Ve18]. Further, it supports the extensibility of the technologies [Wo18].

The increasing popularity of BC/DLT and the associated increased utilization make scalability important [Cr16]. Today for example in the financial economy BC/DLT-systems are incapable of providing the same performance as traditional payment services [Xu17]. Thus scalability is a challenge to a wide usage of BC/DLT [Zh17].

In table 1 criteria for blockchain comparisons identified or used by several authors [BM17; Di17; Hi17; Ka18; Ra18; TT17; VS17; YW18; Zh17] are listed. The work of [TT17] focused primarily on the development of comparison criteria. It proposes a detailed collection of BC/DLT comparison criteria as well as possible expressions of those criteria, that could be selected.

Unlike that [Zh17], [Ra18], [VS17] and [Hi17] do not only describe comparison criteria, but also use them to explicitly make a comparison between BC/DLTs.

Key characteristics of BC/DLTs have been stated by [Zh17] and BC/DLT properties have been used to compare the BC/DLT types public, private and consortium. Further, comparison of six BC/DLT projects including an extensive description of the comparative criteria used is given by [Ra18].[VS17] made a comparison between three BC/DLTs building upon six comparison criteria. And the comparison of seven BC/DLT implementations based on various criteria is described by [Hi17].

Whereas [BM17], [Di17], [YW18] and [Ka18] are using several criteria implicitly in their papers. By [BM17] a comparison of two projects using Ethereum is made. [Di17] describe a framework to analyze private BC/DLTs. Thereby they use performance metrics to evaluate BC/DLTs, which also serve as comparison criteria. A BC/DLT reference model using a layer structure is proposed by [YW18]. It contains several layers that can be used comparatively. Further, this paper describes some BC/DLT application scenarios. A brief comparison of six BC/DLTs is given by [Ka18], taking several aspects of BC/DLT into account.

In order to gain an overview of the comparison criteria, four categories are proposed:

- Criteria to compare **general characteristica** of BC/DLTs and criteria that do not fit into the more detailed categories, e.g. the purpose of a BC/DLT and its reward system (see Table 1a)
- Criteria to compare **security and privacy issues** of BC/DLTs (see Table 1b)

Tab. 1: Criteria for BC/DLT comparison mentioned in literature

(a) General Criteria

Criterion	Mentioned by
Purpose of the BC/DLT	[Hi17], [BM17], [VS17]
Data stored on-chain	[Hi17]
Native Asset offered?	[TT17], [Hi17], [Ra18], [VS17]
Private/Public/Consortium & Permissioned/Permissionless	[TT17], [Hi17], [VS17], [Ka18], [Zh17]
Consensus Model	[Hi17], [VS17], [Ka18], [Zh17], [TT17]
Smart Contracts enabled?	[YW18], [Ka18], [VS17]
Incentive Layer/Reward System	[TT17], [YW18], [Ra18]
Codebase Creation	[Ra18]
Rule Initiation	[Ra18]
Protocol Governance	[Ra18]
Protocol Change	[Ra18]
Data Broadcast	[Ra18]
Transaction Initiation	[Ra18]
Input	[Ra18]
Programmatically Initiated Transactions	[Ra18]
Locus of Execution	[Ra18]
Reference	[Ra18]
Gossiping	[TT17]
Finality	[TT17]
Header Data Structure	[TT17]
Transaction Model	[TT17]
Server Storage	[TT17]
Block Storage	[TT17]
Tokenisation	[TT17]
Asset Supply Management	[TT17]
Interoperability	[TT17]
Intraoperability	[TT17]
Fee System	[TT17]

(b) Criteria Concerning Security and Privacy

Criterion	Mentioned by
Transparency of Decision Making	[Hi17]
Public Key Infrastructure Used?	[Hi17]
Public Key Infrastructure Managing Authority	[Hi17]
Consensus Mechanism	[Hi17], [Ka18], [TT17], [Zh17], [Ra18], [VS17]
(Limits to) Scalability	[Di17], [TT17]
Fault Tolerance	[Di17]
Immutability	[Zh17]
Data Encryption	[TT17]
Data Privacy	[TT17]
Identity Layer	[TT17]
Registration Authority	[Hi17], [Ka18], [VS17], [TT17]

(c) Criteria Concerning Programming

Criterion	Mentioned by
Scripting Language	[TT17], [Hi17], [VS17]
Coding Language	[TT17]
Code License	[TT17], [Ka18], [Ra18], [Hi17]
Software Architecture	[TT17]

(d) Measurable Criteria

Criterion	Mentioned by
Block Release Time	[Hi17]
Transaction Size	[Hi17]
Transaction Rate/Throughput	[Hi17], [Di17], [Zh17]
Latency	[Di17], [TT17]

- Criteria concerning the **programming** of BC/DLTs, e.g. licenses and programming languages (see Table 1c)
- Criteria that focus on **measurable components** of BC/DLTs (see Table 1d)

Table 1 shows, that already a variety of comparison criteria exists. Taking this extensive list as first basis for a BC/DLT comparison reveals difficulties. Finding data on the listed criteria regarding some BC/DLTs may be difficult. And the number of criteria may become an obstacle for a first evaluation of a system's suitability for an own application.

That is why the list of criteria was shortened. As reducing rule is set that at least one criterion of each category should be represented. Thereby it is ensured that the main aspects of BC/DLT are considered. Further, a reduced set of criteria should include those criteria, that seem to be the most important ones. The number of authors mentioning a criterion might be understood as a parameter for importance. That is why a reference of minimum of three authors should be given for a criterion to be selected.

By this constraints the extensive list of criteria shown in table 1 is reduced to the set of criteria shown in table 2.

Tab. 2: Reduced set of comparison criteria

Criterion
BC/DLT Purpose
Native Asset
Private/Public/Consortium; Permissioned/Permissionless
Consensus Model
Smart Contracts
Reward System/Incentive Layer
Consensus Mechanism
Central Registration Authority
Scripting Language
Code License
Transaction Rate/Throughput

This set of criteria is lacking comparison criteria regarding the interoperability and scalability of BC/DLTs. This is caused by the low number of such criteria in the extensive criteria list in table 1. Only three criteria have this focus. The criteria *Interoperability* and *Intraoperability* are only mentioned by [TT17] and the criterion of *(Limits to) Scalability* is only mentioned by two authors [Di17; TT17].

The criteria of the category *measurable* could be understood as criteria for scalability, but even then only one of these criteria meets the condition of being mentioned by at least three authors. Thus the scalability and interoperability are not sufficiently taken into account by the shortened set of criteria in table 2.

4 Comparison of popular BC/DLTs

Nevertheless, this set of criteria has been used to make an exemplary comparison of some popular BC/DLTs. In the comparison selection, blockchains (Bitcoin, Ethereum, Ripple), as well as DLTs (Hyperledger Fabric, Corda), are included. Thereby it is shown that the results are not only valid for either blockchains or distributed ledger technologies. The results of that comparison are shown in table 3.

By that table 3 it is demonstrated, that data regarding nearly every criterion for those blockchains can be found. Thus the used set is a set of criteria, that is well suited for a basic technical blockchain comparison. It serves as a research starting point and first decision support about the blockchain selection. Based on such a comparison a first selection of blockchains to consider for ones use case can be done. The remaining blockchains should then be analyzed using an extended list of criteria. The analysis should then be focused on a custom set of criteria, that seem to be important for each use case.

Further due to the list of criteria it has been shown that criteria concerning scalability and interoperability have not been considered sufficiently. Taking into account the aspect of interoperability interesting can be the level on which interoperability is enabled. Therefore three criteria could be distinguished. A first criterion is the ability to send transactions between different systems. This could be for example sending data from external systems into a BC/DLT system like described by [Ra18]. A second criterion is the possibility to transfer contracts between different systems while retaining the contracts semantics [HLP18]. A third criterion is the extent of interoperability, that has been considered in the systems design phase. Concerning scalability the measurable criteria (see table 1) as a systems possibility to grow are interesting. As a further criterion, especially regarding future trends, can be seen the ability of BC/DLT-systems at high load to enable an enrichment of own capabilities by enabling the usage of capacities of other BC/DLT-systems.

Tab. 3: Comparison of the BC/DLTs Bitcoin, Ethereum, Hyperledger Fabric, Ripple and Corda regarding the shown set of criteria

Criterion	**Bitcoin**	**Ethereum**	**Hyperledger Fabric**	**Ripple**	**Corda**
Purpose	Cryptocurrency [Hi17]	Run Smart Contracts [Hi17]	(Cross-) Industry Use Cases [Hi17], [Ka18], [SSS17]	Global Cross Border Payments [Rid]	Internet-based Management and Automation of Real-world Agreements [Br]
Native Asset	BTC [Hi17], [Bia]	Ether [Etc]	None [Hi17]	XRP [Rif]	None [VS17]
Public/ Private/ Consortium & Permissioned/Permissionless	Public, Permissionless [Bib]	Public [TT17]	Private or Consortium, permissioned [VS17], [TT17]	Public, Permissioned [TT17]	Private, Permissioned [Kh17]
Consensus Model	Transaction Level (blocks and transactions verified) [Hi17], [Ka18]	Ledger Level (blocks and transactions verified) [Hi17], [Ka18]	Transaction Level (pluggable) [Hi17], [Ka18]	Ledger, Transaction Level [Ka18]	Transaction Level (pluggable) [Br]
Smart Contracts	Yes [BT]	Yes [VS17]	Yes [VS17]	Yes [Rie]	Yes [Kh17], [Co]
Reward System/ Incentive Layer	Block Reward [Bic]	Block Reward [Etc]	None	Extrinsic Incentive [Ra18]	None
Consensus Mechanism	Proof-of-Work [Bic]	Proof-of-Work [Etc]	Practical Byzantine Fault Tolerance [Zh17]	XRP Ledger Consensus Protocol [Rib]	pluggable [Co]
Central Registration Authority	None [Hi17]	None [Hi17]	Individually pluggable for each network [Hi17]	RippleNet [Ric]	Individual CA of each network [Li]
Sripting Language	Script [Hi17]	Solidity, Serpent, LLL [Eta]	Golang, node.js, Java [Hya]	-	Any, that targets Java Virtual Machine [R3a]
Code License	Open-Source [Bia]	Open-Source [Etb]	Open-Source [Hyb]	Open-Source [Ria]	Open-Source [Co]
Transaction Rate	7 tx/sec [Hi17]	theoretically no max [Hi17]	>10.000 tx/-sec [Hi17]	1500/sec [Rif]	up to 1000 tx/-sec [R3b]

5 Conclusion

This paper has given a short overview of the importance of comparison criteria for BC/DLT-systems. Currently used comparison criteria in literature have been listed. Then a shortened list of criteria has been narrowed down, which shall serve as a starting point for BC/DLT comparisons. This set has been used in a comparison of five popular BC/DLTs to show, that the data required for a comparison based on the shortened set of criteria is available for some popular BC/DLTs. The literature surveyed has shown that previous research has created only a low number of criteria regarding scalability and interoperability of BC/DLTs.

This paper proposed approaches for criteria on scalability and interoperability. This should further be addressed by future research. In terms of interoperability interesting questions could be related to a system's scope: Can the BC/DLT system operate with other systems? If so, is it interoperable with all other BC/DLTs or only with BC/DLTs of the same type (private/public or permissioned/permissionless) or only with special other BC/DLTs? Further, the level of interoperability will be interesting concerning whether all elements of the BC/DLTs are inter-operable or for example only transactions or only smart contracts.

Further research is required on the development of criteria concerning economical aspects of BC/DLT-systems to extend the range of comparison. An area of interest will also be the impact of economical aspects on scalability and interoperability.

Bibliography

[Bia] Bitcoin Project: Bitcoin - Open Source P2P Money, URL: https://bitcoin.org/en/, Stand: 09. 04. 2019.

[Bib] Bitcoin Project: Einige Dinge, die Sie wissen müssen - Bitcoin, URL: https://bitcoin.org/de/das-sollten-sie-wissen, Stand: 22. 03. 2019.

[Bic] Bitcoin Project: FAQ - Bitcoin, URL: https://bitcoin.org/en/faq#who-controls-the-bitcoin-network, Stand: 09. 04. 2019.

[BM17] Butgereit, L.; Martinus, C.: A Comparison of Two Blockchain Architectures for Inspiring Corporate Excellence in South Africa. In: 2017 Conference on Information Communication Technology and Society (ICTAS). 2017 Conference on Information Communication Technology and Society (ICTAS). IEEE, Durban, South Africa, S. 1–6, März 2017, ISBN: 978-1-4673-8996-9, URL: http://ieeexplore.ieee.org/document/7920656/, Stand: 07. 04. 2019.

[Br] Brown, R. G.: The Corda Platform: An Introduction./, S. 21.

[BT] BTC Inc.: Yes, Bitcoin Can Do Smart Contracts and Particl Demonstrates How, URL: https://bitcoinmagazine.com/articles/yes-bitcoin-can-do-smart-contracts-and-particl-demonstrates-how/, Stand: 09. 04. 2019.

[Bu19] Bundesamt für Sicherheit in der Informationstechnik (BSI), Hrsg.: Blockchain sicher gestalten - Konzepte, Anforderungen, Bewertungen, 2019.

[Co] Corda: Corda | Technology, URL: https://www.corda.net/discover/technology.html, Stand: 09. 04. 2019.

[Cr16] Croman, K.; Decker, C.; Eyal, I.; Gencer, A. E.; Juels, A.; Kosba, A.; Miller, A.; Saxena, P.; Shi, E.; Gün Sirer, E.; Song, D.; Wattenhofer, R.: On Scaling Decentralized Blockchains: (A Position Paper). In (Clark, J.; Meiklejohn, S.; Ryan, P. Y.; Wallach, D.; Brenner, M.; Rohloff, K., Hrsg.): Financial Cryptography and Data Security. Bd. 9604, Springer Berlin Heidelberg, Berlin, Heidelberg, S. 106–125, 2016, ISBN: 978-3-662-53356-7 978-3-662-53357-4, URL: http://link.springer.com/10.1007/978-3-662-53357-4_8, Stand: 19. 06. 2019.

[CWB18] Cooley, R.; Wolf, S.; Borowczak, M.: Blockchain-Based Election Infrastructures. In: 2018 IEEE International Smart Cities Conference (ISC2). 2018 IEEE International Smart Cities Conference (ISC2). IEEE, Kansas City, MO, USA, S. 1–4, Sep. 2018, ISBN: 978-1-5386-5959-5, URL: https://ieeexplore.ieee.org/document/8656988/, Stand: 19. 03. 2019.

[DCP18] Dasaklis, T. K.; Casino, F.; Patsakis, C.: Blockchain Meets Smart Health: Towards Next Generation Healthcare Services. In: 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA). 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA). IEEE, Zakynthos, Greece, S. 1–8, Juli 2018, ISBN: 978-1-5386-8161-9, URL: https://ieeexplore.ieee.org/document/8633601/, Stand: 19. 03. 2019.

[Di17] Dinh, T. T. A.; Wang, J.; Chen, G.; Liu, R.; Ooi, B. C.; Tan, K.-L.: BLOCKBENCH: A Framework for Analyzing Private Blockchains. In: Proceedings of the 2017 ACM International Conference on Management of Data - SIGMOD '17. The 2017 ACM International Conference. ACM Press, Chicago, Illinois, USA, S. 1085–1100, 2017, ISBN: 978-1-4503-4197-4, URL: http://dl.acm.org/citation.cfm?doid=3035918.3064033, Stand: 07. 04. 2019.

[En19] Engelschall, R. S.: Blockchain: Suchen wir nur das Problem zur Lösung? Informatik Spektrum/, 14. Juni 2019, ISSN: 0170-6012, 1432-122X, URL: http://link.springer.com/10.1007/s00287-019-01181-2, Stand: 19. 06. 2019.

[Eta] Ethereum community: Contracts — Ethereum Homestead 0.1 Documentation, URL: http://www.ethdocs.org/en/latest/contracts-and-transactions/contracts.html, Stand: 09. 04. 2019.

[Etb] Ethereum community: What Is Ethereum? - Ethereum Homestead 0.1 Documentation, URL: http://www.ethdocs.org/en/latest/introduction/what-is-ethereum.html, Stand: 09. 04. 2019.

[Etc] Ethereum Foundation: What Is Ether, URL: https://www.ethereum.org/ether, Stand: 22. 03. 2019.

[FF18] Fernandez-Carames, T. M.; Fraga-Lamas, P.: A Review on the Use of Blockchain for the Internet of Things. IEEE Access 6/, S. 32979–33001, 2018, ISSN: 2169-3536, URL: https://ieeexplore.ieee.org/document/8370027/, Stand: 19. 03. 2019.

[Hi17] Hintzman, Z.: Comparing Blockchain Implementations, 2017, URL: https://www.nctatechnicalpapers.com/Paper/2017/2017-comparing-blockchain-implementations/download, Stand: 18. 03. 2019.

[HLP18] Hardjono, T.; Lipton, A.; Pentland, A.: Towards a Design Philosophy for Interoperable Blockchain Systems./, 15. Mai 2018, arXiv: 1805.05934 [cs], URL: http://arxiv.org/abs/1805.05934, Stand: 29. 04. 2019.

[Hu16] Huckle, S.; Bhattacharya, R.; White, M.; Beloff, N.: Internet of Things, Blockchain and Shared Economy Applications. Procedia Computer Science 98/, S. 461–466, 2016, ISSN: 18770509, URL: https://linkinghub.elsevier.com/retrieve/pii/S1877050916322190, Stand: 12. 12. 2018.

[Hya] Hyperledger: Frequently Asked Questions — Hyperledger-Fabricdocs Master Documentation, URL: https://hyperledger-fabric.readthedocs.io/en/release-1.4/Fabric-FAQ.html#chaincode-smart-contracts-and-digital-assets, Stand: 09. 04. 2019.

[Hyb] Hyperledger: Introduction — Hyperledger-Fabricdocs Master Documentation, URL: https://hyperledger-fabric.readthedocs.io/en/release-1.4/whatis.html, Stand: 09. 04. 2019.

[In17] International Standardization Organization: ISO/IEC 30182:2017(en), Smart city concept model — Guidance for establishing a model for data interoperability, 2017, URL: https://www.iso.org/obp/ui/#iso:std:iso-iec:30182:ed-1:v1:en:term:2.5.

[Ka18] Kadam, S.: Review of Distributed Ledgers: The Technological Advances behind Cryptocurrency. In. International Conference Advances in Computer Technology and Management (ICACTM). März 2018, URL: https://www.researchgate.net/publication/323628539_Review_of_Distributed_Ledgers_The_technological_Advances_behind_cryptocurrency, Stand: 07. 04. 2019.

[Kh17] Khan, C.; Lewis, A.; Rutland, E.; Wan, C.; Rutter, K.; Thompson, C.: A Distributed-Ledger Consortium Model for Collaborative Innovation. Computer 50/9, S. 29–37, 2017, ISSN: 0018-9162, URL: http://ieeexplore.ieee.org/document/8048650/, Stand: 09. 04. 2019.

[Ku18] Kumar, T.; Ramani, V.; Ahmad, I.; Braeken, A.; Harjula, E.; Ylianttila, M.: Blockchain Utilization in Healthcare: Key Requirements and Challenges. In: 2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom). 2018 IEEE 20th International Conference on E-Health Networking, Applications and Services (Healthcom). IEEE, Ostrava, S. 1–7, Sep. 2018, ISBN: 978-1-5386-4294-8, URL: https://ieeexplore.ieee.org/document/8531136/, Stand: 19. 03. 2019.

[KV18] Kshetri, N.; Voas, J.: Blockchain-Enabled E-Voting. IEEE Software 35/4, S. 95–99, Juli 2018, ISSN: 0740-7459, 1937-4194, URL: https://ieeexplore.ieee.org/document/8405627/, Stand: 19. 03. 2019.

[La18] Lamberti, F.; Gatteschi, V.; Demartini, C.; Pelissier, M.; Gomez, A.; Santamaria, V.: Blockchains Can Work for Car Insurance: Using Smart Contracts and Sensors to Provide On-Demand Coverage. IEEE Consumer Electronics Magazine 7/4, S. 72–81, Juli 2018, ISSN: 2162-2248, 2162-2256, URL: https://ieeexplore.ieee.org/document/8386868/, Stand: 11. 04. 2019.

[Li] Limited, R.: Joining an Existing Compatibility Zone — R3 Corda Master Documentation, URL: https://docs.corda.net/joining-a-compatibility-zone.html, Stand: 09. 04. 2019.

[Li17] Li, X.; Jiang, P.; Chen, T.; Luo, X.; Wen, Q.: A Survey on the Security of Blockchain Systems. Future Generation Computer Systems/, Aug. 2017, ISSN: 0167739X, URL: https://linkinghub.elsevier.com/retrieve/pii/S0167739X17318332, Stand: 29. 04. 2019.

[Me16] Mettler, M.: Blockchain Technology in Healthcare: The Revolution Starts Here. In: 2016 IEEE 18th International Conference on E-Health Networking, Applications and Services (Healthcom). 2016 IEEE 18th International Conference on E-Health Networking, Applications and Services (Healthcom). IEEE, Munich, Germany, S. 1–3, Sep. 2016, ISBN: 978-1-5090-3370-6, URL: http://ieeexplore.ieee.org/document/7749510/, Stand: 19. 03. 2019.

[Na09] Nakamoto, S.: Bitcoin: A Peer-to-Peer Electronic Cash System./, 2009, URL: https://bitcoin.org/bitcoin.pdf, Stand: 03. 08. 2017.

[PKS16] Pustisek, M.; Kos, A.; Sedlar, U.: Blockchain Based Autonomous Selection of Electric Vehicle Charging Station. In: 2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI). 2016 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI). IEEE, Beijing, S. 217–222, Okt. 2016, ISBN: 978-1-5090-5952-2, URL: http://ieeexplore.ieee.org/document/8281203/, Stand: 11. 04. 2019.

[R3a] R3 Limited: Data Model — R3 Corda Latest Documentation, URL: https://docs.corda.net/releases/release-M7.0/data-model.html, Stand: 09. 04. 2019.

[R3b] R3 Limited: Sizing and Performance — Corda Enterprise Corda Enterprise 3.3 Documentation, URL: https://docs.corda.r3.com/sizing-and-performance.html, Stand: 09. 04. 2019.

[Ra18] Rauchs, M.; Glidden, A.; Gordon, B.; Pieters, G. C.; Recanatini, M.; Rostand, F.; Vagneur, K.; Zhang, B. Z.: Distributed Ledger Technology Systems: A Conceptual Framework. SSRN Electronic Journal/, 2018, ISSN: 1556-5068, URL: https://www.ssrn.com/abstract=3230013, Stand: 07. 04. 2019.

[Ria] Ripple: Home - XRP Ledger Dev Portal, URL: https://developers.ripple.com/, Stand: 09. 04. 2019.

[Rib] Ripple: Introduction to Consensus, URL: https://developers.ripple.com/intro-to-consensus.html, Stand: 09. 04. 2019.

[Ric] Ripple: Join the Network, URL: https://ripple.com/ripplenet/join-the-network/, Stand: 09. 04. 2019.

[Rid] Ripple: Solutions Overview, URL: https://ripple.com/files/ripple_solutions_guide.pdf, Stand: 21. 03. 2019.

[Rie] Ripple: Use Escrows - XRP Ledger Dev Portal, URL: https://developers.ripple.com/use-escrows.html, Stand: 09. 04. 2019.

[Rif] Ripple: XRP, URL: https://ripple.com/xrp/, Stand: 22. 03. 2019.

[SSS17] Sankar, L. S.; Sindhu, M.; Sethumadhavan, M.: Survey of Consensus Protocols on Blockchain Applications. In: 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, Coimbatore, India, S. 1–5, Jan. 2017, ISBN: 978-1-5090-4559-4, URL: http://ieeexplore.ieee.org/document/8014672/, Stand: 09. 04. 2019.

[Te] Tendermint Inc.: Cosmos Network - Internet of Blockchains, URL: https://cosmos.network/whitepaper, Stand: 29. 04. 2019.

[TS] Thomas, S.; Schwartz, E.: A Protocol for Interledger Payments./, URL: https://interledger.org/interledger.pdf, Stand: 29. 04. 2019.

[TT17] Tasca, P.; Tessone, C. J.: Taxonomy of Blockchain Technologies. Principles of Identification and Classification./, 31. Mai 2017, arXiv: 1708.04872 [cs], URL: http://arxiv.org/abs/1708.04872, Stand: 02. 08. 2018.

[va] vasa: ContractPedia: An Encyclopedia of 40+ Smart Contract Platforms, URL: https://hackernoon.com/contractpedia-an-encyclopedia-of-40-smart-contract-platforms-4867f66da1e5, Stand: 06. 04. 2019.

[Ve18] Verdian, G.; Tasca, P.; Paterson, C.; Mondelli, G.: Quant Overledger Whitepaper, 2018, URL: https://www.quant.network/wp-content/uploads/2018/09/Quant_Overledger_Whitepaper-Sep.pdf, Stand: 29. 04. 2019.

[VS17] Valenta, M.; Sandner, P.: FSBC Working Paper - Comparison of Ethereum, Hyperledger Fabric and Corda, Juni 2017, URL: http://explore-ip.com/2017_Comparison-of-Ethereum-Hyperledger-Corda.pdf, Stand: 22. 03. 2019.

[Wo] Wood, D. G.: POLKADOT: VISION FOR A HETEROGENEOUS MULTI-CHAIN FRAMEWORK./, URL: https://polkadot.network/PolkaDotPaper.pdf, Stand: 29. 04. 2019.

[Wo18] Wood, D.: A Future History of International Blockchain Standards. The Journal of the British Blockchain Association 1/1, S. 1–10, 4. Juli 2018, ISSN: 25163949, 25163957, URL: https://jbba.scholasticahq.com/article/3724-a-future-history-of-international-blockchain-standards, Stand: 19. 06. 2019.

[Xu17] Xu, X.; Weber, I.; Staples, M.; Zhu, L.; Bosch, J.; Bass, L.; Pautasso, C.; Rimba, P.: A Taxonomy of Blockchain-Based Systems for Architecture Design. In: 2017 IEEE International Conference on Software Architecture (ICSA). 2017 IEEE International Conference on Software Architecture (ICSA). IEEE, Gothenburg, Sweden, S. 243–252, Apr. 2017, ISBN: 978-1-5090-5729-0, URL: http://ieeexplore.ieee.org/document/7930224/, Stand: 29. 04. 2019.

[YW18] Yuan, Y.; Wang, F.-Y.: Blockchain and Cryptocurrencies: Model, Techniques, and Applications. IEEE Transactions on Systems, Man, and Cybernetics: Systems 48/9, S. 1421–1428, Sep. 2018, ISSN: 2168-2216, 2168-2232, URL: https://ieeexplore.ieee.org/document/8419306/, Stand: 07. 04. 2019.

[Zh17] Zheng, Z.; Xie, S.; Dai, H.; Chen, X.; Wang, H.: An Overview of Blockchain Technology: Architecture, Consensus, and Future Trends. In: 2017 IEEE International Congress on Big Data (BigData Congress). 2017 IEEE International Congress on Big Data (BigData Congress). IEEE, Honolulu, HI, USA, S. 557–564, Juni 2017, ISBN: 978-1-5386-1996-4, URL: http://ieeexplore.ieee.org/document/8029379/, Stand: 02. 04. 2019.

HomeCA: Scalable Secure IoT Network Integration

Robert Müller[1] Corinna Schmitt[2] Daniel Kaiser[3] Marcel Waldvogel[4]

Abstract: Integrating Internet of Things (IoT) devices into an existing network is a nightmare. Minimalistic, unfriendly user interfaces, if any; badly chosen security methods, most notably the defaults; lack of long term security; and bugs or misconfigurations are plentiful. As a result, an increasing number of owners operate unsecure devices.
Our investigations into the root causes of the problems resulted in the development of *Home Certificate Authority* (HomeCA). HomeCA includes a comprehensive set of secure, vendor-independent interoperable practices based on existing protocols and open standards. HomeCA avoids most of the current pitfalls in network integration by design. Long-term protocol security, permission management, and secure usage combined with simplified device integration and secure key updates on ownership acquisition pave the way toward scalable, federated IoT security.

Keywords: Home Certificate Authority (HomeCA); Communications technology; Internet of Things (IoT); Network security; Wireless communication; Wireless Local Area Network (WLAN)

1 Introduction

The Internet of Things (IoT) consists of a myriad of network-enabled devices serving in various parts of our daily life (e.g. business, household, personal health and entertainment), which are connected via the Internet. IoT devices are vulnerable, not least owing to their use of wireless connectivity paired with limited defence resources, which brings an enormous diversity in possible attacks [BW15].
The number of incidents of lost control over IoT devices is on the rise [UC17]. They are then used to mount Distributed Denial of Service (DDoS) attacks [Wa16] or abuse the device itself, risking the owner's privacy and security. The owner will rarely be aware of these abuses, due to a combination of the devices' limited resources and minimal or non-existent user interfaces. Lack of long-term security updates, lack of vulnerability notifications to the users plus often complicated and impractical update procedures further challenge their use for any serious and safe usage.

[1] Universität Konstanz, Department of Computer and Information Science, Distributed Systems Laboratory, Universitätsstraße 10, 78457 Konstanz, Germany robert.mueller@uni-konstanz.de
[2] Universität der Bundeswehr München, Research Institute CODE, Werner-Heisenberg-Weg 39, 85577 Neubiberg, Germany corinna.schmitt@unibw.de
[3] University of Luxembourg, Security and Networking Research Group Netlab, Maison du Nombre 6, Avenue de la Fonte, 4364 Esch-sur-Alzette, Luxembourg daniel.kaiser@uni.lu
[4] Universität Konstanz, Department of Computer and Information Science, Distributed Systems Laboratory, Universitätsstraße 10, 78457 Konstanz, Germany marcel.waldvogel@uni-konstanz.de

doi:10.18420/inf2019_21

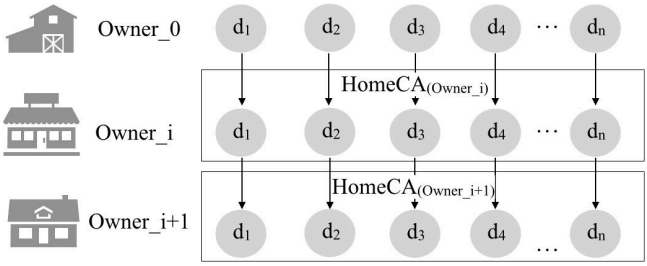


Fig. 1: HomeCA coverage on IoT devices life cycle

To add insult to injury, the humongous problems caused by insecure and sometimes insecureable devices (i.e. without interface/means to install security patches) are complemented by poor pairing protocols, which are insecure by design or implementation: Many application profiles of the ZigBee specification (low-power, low bandwidth, and low range) often used in smart homes allow, by definition, any malicious device to join the network and, thus, become trusted [Zi15]. Wi-Fi is not much better, recommending insecure defaults [VP16] and opening barn doors with Wi-Fi Protected Setup (WPS) [Wi14] which allows offline attacks [Bo14].

The manufacturers have been asked to provide more security by forcing the idea of privacy-by-design on them and applying data protection regulations [Wa16], though they are not always applied due to missing incentives and controls when not following the requests and rules. From our research on the field of wireless security on IoT devices we identified the following four core problems: (1) Vendors deliver their devices with simple instead of secure setup (savings in customer support requests). (2) Vendors do rarely provide security updates, provide no easy way of installing them, remove functionality unnecessarily together with updates [Ha16], or eventually go out of business. (3) Users do often leave passwords at the manufacturer's default settings, if they can change them at all. (4) Devices are often accessible from the Internet, even if it is not required for their functionality.

With our Home Certificate Authority (HomeCA) approach we determine new ways to pair home devices, to get rid of passwords, and to take advantage of new developments such as built-in security keys in micro controllers [Mi18]. At the same time, we are trying to remain vendor neutral and avoid unnecessary trust in the device keys, as manufacturing errors, database leaks, and low entropies have been known to cause problems in other published cases [Ma16].

HomeCA includes a lifecycle consisting of four phases: (1) Manufacturing, (2) Ownership Change, (3) Connection Establishment, and (4) Refresh. Figure 1 illustrates the transitions in the lifecycle of IoT devices from manufacturing, resale, use in a home network to the possibility of a reuse at another home, e.g., a friends home or during vacation.

With the process included in those phases, HomeCA supports secure pairing, and optional gatewaying can be provided, resulting in (1) removing insecure passwords and (2) reducing the dependency on updates, all in a vendor-independent layer with little overhead. HomeCA creates a protocol layer that (3) provides security, (4) does not require complex

and potentially insecure user interaction, and (5) can help reduce the trust required in the manufacturer.
HomeCA is a security protocol layer that covers the IoT devices within a trustworthy home network. It uses a machine-to-machine (M2M) authentication protocol, which allows authentication between devices and applications without user interaction except for an approval step to mitigate automated attacks, for the integration of new IoT devices into the network. It is designed to protect the devices from unauthorized manipulation and access using a Certification Authority (CA) within the home network. Threats that are denied through HomeCA and the used *methods* are: (1) Lacking or corrupted (security) updates *(verification)*, (2) Corruption of the manufacturer itself *(verification)*, (3) Attacks from the Internet aimed at eavesdropping communication *(encryption)*, and (4) Attacks from the Internet aimed at taking control of the IoT device *(certificates)*.
The paper is structured as follows: Section 2 focuses on related work in the area of securing inhomogeneous IoT networks. Section 3 introduces our HomeCA model leading to supported workflows in Section 4 and secure key update in Section 5. A prototype implementation is presented in Section 3.4, and Section 6 concludes the paper.

2 Related Work

Over the last decades many investigations took place to identify challenges in IoT [Su12], especially in the area of security of Public Key Infrastructure (PKI) [GM94], [Sv16] and in the area of key management [VP16], [Sc15].
However, it was shown that the IEEE 802.11 implementation Wi-Fi Protected Access 2 (WPA 2) group keys can be attacked based on the quality of the used random number generator [VP17], which often is embedded in IoT devices.
A good overview and a broad analysis of Authentication and Access Control used in the IoT can be found in [LXC12]. An abstract approach for IoT integration by a central signing authority server is followed in a Patent Application in [Sh18]. Research on secure and resources saving integration of constrained devices into the IoT is presented in [Kl15], with a certificateless signcryption scheme targeting at Wireless Sensor Networks (WSN) rather than Wireless Local Area Network (WLAN) networks. WPS [Wi14] and similar technologies do not provide the possibility to integrate a large number of devices. Focussing on smart homes, security challenges are studied in [BJD16], while in [Si15] a network centric approach is proposed, in which a central entity, remotely similar to HomeCA, is located outside the smart home network that provides services through a web interface to securely control smart home devices in the local network.

3 HomeCA Model

For the design and development of HomeCA we assume the following use case illustrating a life cycle of today's IoT devices: A set of IoT devices $\{d_n, n \in \mathbb{N}\}$ is created at a factory

and being sold at a retail store. Once they are bought, they are integrated and operated in a first home network. Eventually, they are moved to a second home, showing disassociation and re-association of the IoT device.

3.1 Security Model

IoT devices are target of attacks that try to access the sensitive data or try to gain control over the device. Devices taken over are used for illegal actions like Distributed Denial-of-Service (DDoS) attacks. This type of attempts, usually originated from the Internet, trying to access IoT devices within the private network, are detected by HomeCA through their suspicious or blacklisted source IP address and thus are dismissed. There is no unauthorized access to the IoT devices since traffic must pass the check of HomeCA and thus cannot reach the IoT device.
For privacy reasons it shall not be disclosed to the manufacturer, how many devices are within the personal/private network managed by HomeCA. This is achieved by signatures between the manufacturer, devices and HomeCA.
Attackers try to find and access IoT devices connected to the Internet. Methodology and tools to exploit known vulnerabilities of devices and protocols used in the IoT are described in [MM15]. Other information about the devices may be stolen from a manufacturer or service website that the IoT device is eventually connected to, such as a cloud service or a service provider API.
Next, an attacker tries to log in with default credentials to collect all devices that have not changed the factory-default username and password combination. Once successful, an attacker can configure the IoT device as he/she pleases. HomeCA protocol protects the access to IoT devices and therefore prevents this type of direct access to the IoT devices, except if they are legit.
Changes to the IoT device configuration requested by the owner are redirected to the HomeCA web-interface. As a consequence, multiple access attempts for IoT devices within the private network can be detected by the unusual high number within a definite time window. Allowed access attempts are additionally protected against maliciously configured bots by the use of i.e. "Completely Automated Public Turing test to tell Computers and Humans Apart"(CAPTCHA) [Ca17] or similar technologies capable of preventing bots from performing specific operations automatically. This is triggered by exceeding a variable threshold t on the number of devices being integrated at once.
Next, the HomeCA Lifecycle is introduced, which illustrates the required coupling between the HomeCA and an IoT device.

3.2 HomeCA Lifecycle

The HomeCA lifecycle of an IoT device consists of the following four steps, where the last one is optional (for details see Section 4):

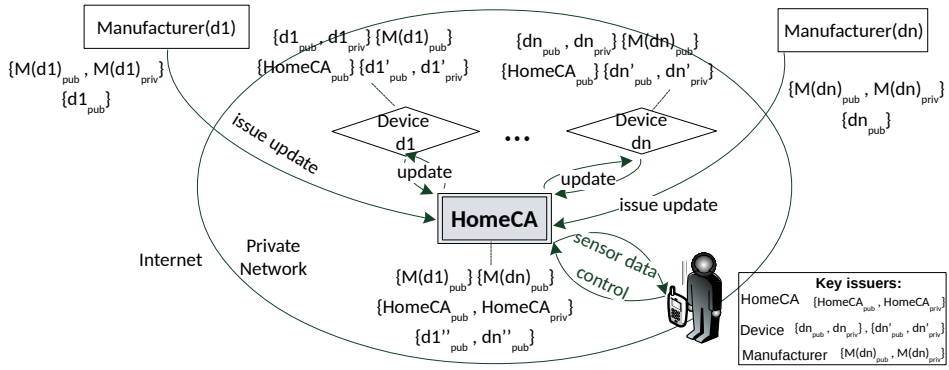


Fig. 2: HomeCA System Overview

1. **Manufacturing:** Initial key pair creation, public key delivery *(optional)*.
2. **Ownership change:** Release *(active or passive)*, HomeCA discovery, key verification *(optional)*, key update, certificate creation and revocation, rights management *(optional)*.
3. **Connection establishment:** Service discovery, Datagram Transport Layer Security (DTLS) connection setup.
4. **Refresh *(optional)*:** Liveness verification, certificate lifetime extension, rights management update.

Figure 2 shows the basic mechanisms of the protocol set-up for the secure communication of IoT devices with their two manufacturers and within a private/personal network.
Initially, keys for encrypted communication are exchanged between the manufacturers (M) $M_{1..n}$ and corresponding devices (d) $d_{1..n}$. Assuming an honest manufacturer for the time during transfer of posession), this offline exchange at production site is considered secure. Second, secure keys for bi-directional communication are exchanged between HomeCA and $d_{1..n}$, namely $d1''_{pub}$ .. dn''_{pub}. This happens within the private/home network and provides a trusted communication channel between the IoT device and HomeCA for access and control of the IoT devices $d_{1..n}$. No user-interaction for execution is required. Afterwards, the initially pre-shared keys are then shared by the devices (d) $d_{1..n}$ with HomeCA to verify an update issued by $M_{1..n}$ before it is applied by $d_{1..n}$.
In order to make the interface scalable for many devices we chose an interface that requires no human interaction except for security approvals when a large number of devices exceeding a variable threshold is handled. The device integration is based on a trust relationship between manufacturer, device and home network. The protocol execution steps for the handover of a device between multiple HomeCA are shown in Figure 3. Willing to be connected with another HomeCA, device d sends a request to the new HomeCA that is available within the network. To switch from the initial owner 0 to owners i and later $i + 1$, the device d asks the current HomeCA to be released. Afterwards d registers with the new HomeCA via a DTLS [RM12] handshake and thus is under new administration.

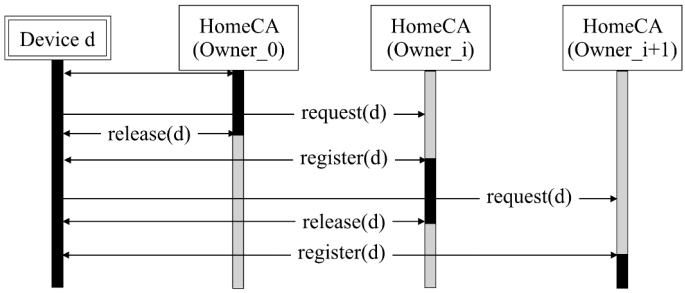


Fig. 3: HomeCA protocol integrating a new device

3.3 Security Use Cases

The use cases that require the secure infrastructure to update its trust relationship are:
Device d is brought into a private wireless network for the first time. HomeCA recognizes the new device and has to authenticate against d to prove it is authorized to manage d. Two ways exist to do the aforementioned: (1) Option one is chosen when the device is directly purchased at the manufacturer M. At purchase, a unique device ID of d that consists of a hash on identifiers e.g., Media Access Control (MAC) address, Bluetooth device identifier, and Personal Area Network Identifier (PANID) is transmitted from M to the buyer which imports them into HomeCA or submits an E-mail address to send the device IDs to. When the device d appears within the HomeCA network, they perform a handshake and can authenticate and establish a secure communication with d. (2) Option two is chosen when d is bought at a retail store. d and HomeCA meet in a private network considered secure when no interference (i.e. more than one HomeCA instance or other connection attempts) is being detected by d. HomeCA is actively asking for a new to be integrated device, when it knows i.e. from the application registry which sensor data to expect from the sensor of the IoT device d. Diffie-Hellman key exchange between d and HomeCA is used to obtain k to create new public and private keys.
Device d is sold /given away /defect: If HomeCA did not see its IoT devices for a period of time t, the certificates for that specific device are removed from HomeCA and must re-authenticate if it appears in the private network again at a later point in time. The initial time window is set to $t = 10$ days but shall be adopted based upon the dynamic of the actual user behaviour of the private network, i.e. the rate at which devices are introduced and removed.
Device software d has a security weakness or needs to be updated/patched: Only HomeCA can approve and sign a trustworthy update. Attackers without this certificate cannot create a valid update for the device. HomeCA monitors and keeps track of all updates on the IoT devices.
If the device d is compromised and behaves strangely, HomeCA removes the certificate for the device d automatically based on a revocation list. The revocation list is signed by HomeCA and provided for all associated devices as well for the scenario of devices

communicating directly with each other.
If the manufacturer M is compromised and no longer trustworthy, HomeCA is able to inform the IoT devices assigned with HomeCA about this situation.

3.4 Model Implementation and First Evaluation

In order to analyse and evaluate HomeCA, we chose to implement a prototype on physical devices. A set of three Raspberry Pi 3 Model B[5] is used to set up a model of one HomeCA and two IoT devices or optionally two HomeCA and one IoT device switching HomeCAs. Java is used as a platform-independtent language.
For network access and discovery purposes we employ a WPA2-Enterprise Radius Server (see Section 4.4). Also, the necessity of a database on the IoT device to maintain the certificates besides the connection information is identified and shall be used in the next release of HomeCA.

4 HomeCA Workflows

HomeCA workflows are used to create a trusted association between the device and the network and thus the other connected devices. Thereby, reasonable but minimal trust is expended on the manufacturer of the device and the HomeCA itself. To achieve this, as little sensitive information as necessary is processed by other devices. Key pairs are be solely managed by the device owning them, meaning private key material is not shared or transmitted unless absolutely necessary.

4.1 Manufacturing: Initial Key Pair Creation

Ideally, the IoT device itself is solely responsible to create the key pair and the private key is not accessible from the outside. This requires a good source of random bits, requiring the devices to be more complex. Key creation at manufacturing time comes with the following three options: Local generation, key feeding, and randomness feeding.
Local generation: This is the ideal, but requires a good internal source of reliable and confidential random bits, together with ample processing power. If the random number generator (RNG) does not meet these criteria, the key will be weak and can be guessed with little effort [Bo14].
Key feeding: Device does not possess a good enough RNG. Therefore, a key pair is created by an external device and fed to the IoT, e.g., during burn-in test. As a result, the external device owned by the manufacturer knows the private key. This key can then be obtained by

[5] Specification available at `https://www.raspberrypi.org/products/raspberry-pi-3-model-b/`

Tab. 1: IoT Device Protection Mechanisms and Access Rights

Service	Allowed function	Required protection	Gateway usage recommended
Heating	Report temperature	TLS 1.2	NO
Garage Door	Open/Close	SSLv3	YES
Key Generator	Online Authentication	TLS 1.3	NO
Entertainment	Volume control	none	YES
SmartHome	Steer Heating, Lights	SSLv2	YES
Coffee Maker	Preheat machine	SSLv2	YES
Fitness	Track steps	TLS 1.0	YES
Health	Blood Sugar monitor	TLS 1.1	YES
SmartWatch	Messages, Notifications	TLS 1.3	NO

third parties [SB15] and therefore should not be considered secret.

Randomness feeding: An alternative way to cope with a bad on-chip RNG is to use an external random source, e.g., provided again during burn-in test. The device then calculates its own key pair based on that data and keeps the private key secret. However, anyone knowing the key generation algorithm (often public) and the random bits fed, can recreate the private key and verify it with the public key, duplicating the problems of *key feeding* above.[6]

HomeCA uses local generation – in our prototype implementation with a hardware random number generator. Even if the key is generated on the device itself, the manufacturer may be able to extract the key using low-level device debug mechanisms such as JTAG [IE13].[7] Table 1 lists exemplary IoT devices we found. The devices are shown with provided functionality, protection mechanism and decision about whether the HomeCA Gateway shall be used based on the required protection. Gateway functionality (cf. Section 4.7) is proposed for some (legacy) devices with protection mechanism considered less secure to decrease the risk of attackers exploiting the communication protocol. This is explicitly used for services that might seem not security relevant, as e.g., with Smart Homes it would be disturbing for the user and fun for a hacker to turn lights and heating on/off.

4.2 Manufacturing/Resale: Public Key Delivery

After the key pair creation, optionally, the public key may be extracted from the device and associated with the device serial number. If this public key remains associated with the particular device during the entire sales chain, it can be used to simplify the ownership change process (cf. Section 4.4).

[6] Mixing in data from a weak local RNG is possible, but will not significantly increase the attack effort.

[7] A particularly malicious manufacturer might even include a fuse bit, which after activation would hide itself and the private key access possibility.

4.3 Ownership Change: Release

A device first must be released by its previous owner, before it starts looking for a new home. This is best done by sending an explicit command to the device by the previous owner (which, on the first sale, is the manufacturer), called an *active release*. Fallback methods (*passive release*) should include a timeout (not having seen the HomeCA for a predefined period, e.g. a week). To some extents, this can be considered a variation of the Resurrecting Duckling security policy model [SA99], which describes a transient ownership relation of a device between multiple owners.

4.4 Ownership Change: HomeCA Discovery & Key Verification

A released IoT device, whether actively or passively, will search for a new home network. After connecting to a network, a device will first use Multicast DNS Service Discovery (mDNS-SD, also known by its implementations Bonjour and Avahi) [CK13] to discover the HomeCA. The question to be answered is how the IoT device connects to the network in the first place. Two possibilities exist:
(1) Probe an open network: This is easiest to implement, but will make it prone to networks set up as traps by adversaries trying i.e. to collect IoT device information.
(2) Connect to a WPA2-Enterprise (WPA2 EAP-TTLS, Extensible Authentication Protocol Tunneled Transport Layer Security Authenticated Protocol Version) network. 802.1X is used to authenticate a client in the network with an individual Master Session Key (MSK) for each session. The authentication phase is extended such that contacting the HomeCA in a limited fashion prior to having full network access is already possible during this phase [Sm12]. Later, connections to a WPA Enterprise network will present the HomeCA certificate to safely connect in EAP-TLS mode.
As long as a device has not been accepted and granted full access by a HomeCA (cf. Section 4.5), it will keep probing.
When a device public key has been delivered as part of the order, it may be pre-configured in the HomeCA, allowing later automatic joining as soon as this device comes within network range. This is of course the most comfortable and thus recommended operation.
Without public key delivery, the device will try to join the network and contact the HomeCA. The device will sit in this *unverified* state until the user has manually confirmed the device addition. In the simplest case, this is performed similar to the WPS Push-Button authentication [Wi14]. However, we envision a control application on the owner's smartphone, which displays device information (e.g., name, type, joining time) to verify it is actually the correct device. This application may then also be used for rights management.

4.5 Ownership Change: Key Update, Rights Management & Certificate Creation

The key currently associated with a device might be known to the manufacturer (cf. Section 4.1) and/or previous owner. Anyone knowing the key can directly impersonate this device or perform an undetectable Man-In-The-Middle (MITM) attack. Thus, reusing the same key is to be avoided. Therefore, on association, a new key is generated. To avoid accidental association with a foreign HomeCA, an initial integration of a device must be accepted within a User Interface (UI) presented to the user, i.e. on a smartphone app. Also, an IoT device will only accept a new HomeCA as its owner, if it has been previously released (cf. Section 4.3). Only if the user actively searches for devices, both known and unknown ones are shown. Otherwise unknown ones are hidden.
On a key update, any sensitive data on a device is also wiped, making them unavailable to the previous owner. This may include access rights to other devices, measurement data, or information received from other devices while with the previous owner [VA14].
Optionally, the owner, possibly through the HomeCA app, can assign rights this device should have toward other devices. This is represented as an Access Control List (ACL) [GPR13, Le84], possibly in matrix form, listing allowed operations per target device or device group as shown in column two, *Allowed function*, in Table 1.
The HomeCA signs the public key together with device information, validity period, and the optional ACL. This certificate is passed to the device.

4.6 Refresh: Liveness Verification

Liveness verification is initiated by the IoT device and used to state their availability to HomeCA, to regularly update the certificate with the connection information and optional ACL. It also allows IoT devices to reactivate the certificate-based relationship with HomeCA after a long period without interaction, e.g., during holidays that exceeds the time window t.

4.7 Service Discovery & Connection Establishment

Configurationless service discovery for HomeCA is done using DNS Service Discovery (DNS-SD) over multicast DNS (mDNS) [CK13]; widely known as *Zeroconf* or *Bonjour*. Because this solution does not protect the users' privacy, we include a privacy extension as presented in [KW14].
An initiator device A presents its certificate implementing a TLS or DTLS profile [TF16] to a target device B (with optional ACL signed by HomeCA), claiming authorization to access the device and perform particular operations.
As a result, each connection is secured by (at least three) layers of additional protection: (1) Device A can only be access other HomeCA controlled devices B when presenting a valid certificate. (2) With ACLs in place, the actual operations that A can perform on

Tab. 2: IoT device ownership change: Key knowledge

Entity	Manufacturer key known	Old device key known	Key delta	New device key known
Manufacturer	YES	NO	NO	NO
Previous owner	NO	PERHAPS	NO	NO
IoT device	YES	YES	YES	YES
New owner	NO	NO	YES	YES

B are positively enumerated. (3) Device *B* will only accept a subset of protocols which are considered secure by HomeCA. (4) The gateway (that could be HomeCA itself or a dedicated device, cf. Table 1) can translate between incompatible security protocols and provide additional content filters for insecure devices.
If gatewaying/proxying is not supported by the particular application layer protocol for the device, it can be emulated by TLS Server Name Indication (SNI) as used in Domain Fronting [Fi15].

5 Secure Key Update

As we have seen, ownership change for IoT devices (Section 3.2) is common, entropy may be hard to come by (Section 4.1), and — as they can work with sensitive data but are hard to control (Section 3.1) — should receive minimal trust.
While open adversities by the manufacturer may be rare, negligence or process errors are not uncommon: In the early days of networking equipment, it started with batches of network adapters with matching hardware addresses. Today, it is identical or low-entropy encryption keys or the stealing of these keys in the facility [SB15]. So, these initial keys should be minimally trusted, but any entropy in them should be kept.
Even though the new private key should only be known to the device, the HomeCA can be sure it includes any additional entropy from the HomeCA as part of the key update process. This is achieved through knowledge splitting (cf. Table 2), i.e. the parties (manufacturers, devices, and owners) only know the keys they are working with, such that no party knows the complete set of cryptographic keys of the communication system.
We will base the description of our secure key update algorithm on the following three entities: (1) Device d with a possibly non-unique ECC (Elliptic Curve Cryptography) key, that might also be known to a third party, i.e. manufacturer M. (2) Manufacturer M that could know the ECC key of d. (3) HomeCA, e.g. running within an owners Smart Home system and is a trustworthy entity that shall not know the key.
Device d is created by manufacturer M. d is flashed with its software during the manufacturing process. At this time, M provides the following keys that are stored on d: A public key $\{M_{pub}\}$ for the communication with M, and a unique public key infrastructure (PKI) key pair $\{d_{pub}, d_{priv}\}$. M stores $\{d_{pub}\}$ only but must be expected to know the private key, too.

HomeCA can verify for d, that update U is actually from M since the update is signed with the certificate of M. Access to d is only granted via HomeCA because device d only accepts messages encrypted with $\{d_{pub}\}$ which is known to HomeCA only. The protocol steps for the integration of d into the home network with HomeCA protocol are the following:
(1) A key pair k for device d including entropy from the previous key and additional entropy from the information exchange between the HomeCA and the device is created. (2) HomeCA continuously broadcasts its presence within the private network and requests new devices for authentication. (3) d authenticates against HomeCA with its key pub_d . (4) Auxiliary condition: Geographically restricted and within a small time window to reduce the attack vector. (5) d and HomeCA create a common Diffie-Hellman Key. The result is not predictable by both. (6) Entropy check: Verification of the quality of the random number r. (7) d creates a new key pub_{d2} and $priv_{d2}$, that are obtained by multiplication of pub_d and $priv_d$ with k. (8) HomeCA signs the new key pub_{d2}, that HomeCA can obtain from $k * pub_d$ and provides the certificate to d.
As a result the private key $priv_d$ of d is known to d and potentially also to the the manufacturer M. The key k is known to d and HomeCA, whereas $priv_{d2} = k * priv_d$ is only known to d. At the same time, HomeCA knows, that d has integrated material.

6 Conclusions and Future Work

This paper addresses the challenges of the integration of IoT devices into a private network. We present our mostly automated security protocol HomeCA. It focuses mainly on two functions: First, secure integration, which relies on PKI cryptography that prevents attacks from the Internet. Second, scalability to a large number of IoT devices, by the automated integration process without user interaction based on defined protocol steps within the private network.
The steps described in Section 4 ensure that on first purchase and later ownership changes, the keys are updated securely, even when the device lacks a reliable entropy source. The processes are designed to ensure long-term compatibility and security, even when it is expected that the devices will not be provided with security updates.
For the future it is envisioned to extend the implementation of the HomeCA protocol in an environment with more devices to study and analyze its viability without requiring significant changes on existing IoT devices. This full implementation will also allow deployment to verify several parameters, including the validity periods and timeouts.

Bibliography

[BJD16] Bugeja, Joseph; Jacobsson, Andreas; Davidsson, Paul: On Privacy and Security Challenges in Smart Connected Homes. In: European Intelligence and Security Informatics Conference (EISIC). IEEE, 2016.

[Bo14] Bongard, Dominique: , Offline bruteforce attack on WiFi Protected Setup. `http://archive.hack.lu/2014/Hacklu2014_offline_bruteforce_attack_on_wps.pdf`, 2014. (Accessed: 2019/06/17).

[BW15] Barcena, Mario Ballano; Wueest, Candid: , Symantec Secrity Response: Insecurity in the Internet of Things. https://tinyurl.com/yb227t7d, 2015. (Accessed: 2019/06/17).

[Ca17] Carnegie Mellon University: , CAPTCHA: Telling Humans and Computers Apart Automatically. http://www.captcha.net/, 2017. (Accessed: 2019/06/17).

[CK13] Cheshire, S.; Krochmal, M.: , IETF RFC 6763 DNS-Based Service Discovery. https://tools.ietf.org/html/rfc6763, 2013. (Accessed: 2019/06/17).

[Fi15] Fifield, David; Lan, Chang; Hynes, Rod; Wegmann, Percy; Paxson, Vern: Blocking-resistant communication through domain fronting. In: Proceedings on Privacy Enhancing Technologies. pp. 46–64, June 2015.

[GM94] Gander, Martin J.; Maurer, Ueli M.: On the Secret-Key Rate of Binary Random Variables. In: International Symposium on Information Theory. IEEE, 1994.

[GPR13] Gusmeroli, Sergio; Piccione, Salvatore; Rotondi, Domenico: A capability-based security approach to manage access control in the Internet of Things. Mathematical and Computer Modelling (Elsevier), 58:1189–1205, 2013.

[Ha16] Harmon, Elliot: , Don't Hide DRM in a Security Update. https://www.eff.org/deeplinks/2016/09/dont-hide-drm-security-update, 2016. (Accessed: 2019/06/17).

[IE13] IEEE Computer Society: 1149.1-2013 - IEEE Standard for Test Access Port and Boundary-Scan Architecture. In: Working Group: Boundary Scan Architecture - Standard Test Access and Boundary Scan Architecture WG P1149.1. 2013.

[Kl15] Klugah-Brown, Benjamin; Aristotle, John Bosco; Ansuura, Kanpogninge; Qi, Xia: A Signcryption Scheme from Certificateless to Identity-based Environment for WSNs into IoT. International Journal of Computer Applications (0975 – 8887), 120(9), 2015.

[KW14] Kaiser, Daniel; Waldvogel, Marcel: Efficient Privacy Preserving Multicast DNS Service Discovery. In: Workshop on Privacy-Preserving Cyberspace Safety and Security (IEEE CSS). IEEE, pp. 1229–1236, 2014.

[Le84] Levy, Henry M.: Capability-Based Computer Systems. Butterworth-Heinemann, 1984.

[LXC12] Liu, Jing; Xiao, Yang; Chen, C. L. Philip: Authentication and Access Control in the Internet of Things. In: 32nd International Conference on Distributed Computing Systems Workshops. 2012.

[Ma16] Maire O'Neill: Insecurity by Design: Today's IoT Device Security Problem. Engineering (Elsevier), 2:48–49, 2016.

[Mi18] Microchip Technology Inc.: , AWS Zero Touch Secure Provisioning Platform. http://www.atmel.com/applications/iot/aws-zero-touch-secure-provisioning-platform/default.aspx, 2018. (Accessed: 2019/06/17).

[MM15] Markowsky, Linda; Markowsky, George: Scanning for Vulnerable Devices in the Internet of Things. In: 8th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications. 2015.

[RM12] Rescorla, E.; Modadugu, N.: , IETF RFC 6347 Datagram Transport Layer Security Version 1.2. https://tools.ietf.org/html/rfc6347, 2012. (Accessed: 2019/06/17).

[SA99] Stajano, Frank; Anderson, Ross: The Resurrecting Duckling: Security Issues for Ad-hoc Wireless Networks. In: International Workshop on Security Protocols. April 1999.

[SB15] Scahill, Jeremy; Begley, Josh: The Great SIM Heist. The Intercept, February 2015. https://theintercept.com/2015/02/19/great-sim-heist/ (Accessed: 2019/06/17).

[Sc15] Sciancalepore, Savio; Capossele, Angelo; Piro, Giuseppe; Boggia, Gennaro; Bianchi, Giuseppe: Key Management Protocol with Implicit Certificates for IoT systems. In: 1st International Workshop on IoT Challenges in Mobile and Industrial Systems. 2015.

[Sh18] Shah, Rashmikant B; Weis, Brian E; Kumar, Kannan; Nayak, Manoj Kumar: , Zero-touch iot device provisioning, November 1 2018. US Patent App. 15/582,113.

[Si15] Sivaraman, Vijay; Gharakheili, Hassan Habibi; Vishwanath, Arun; Boreli, Roksana; Mehani, Olivier: Network-Level Security and Privacy Control for Smart-Home IoT Devices. In: 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). IEEE, pp. 163–167, Oct 2015.

[Sm12] Smith, Tim: , 802.11 Sniffer Capture Analysis - WPA/WPA2 with PSK or EAP. https://supportforums.cisco.com/t5/wireless-mobility-documents/802-11-sniffer-capture-analysis-wpa-wpa2-with-psk-or-eap/ta-p/3116990, 2012. (Accessed: 2019/06/17).

[Su12] Suo, Hui; Wan, Jiafu; Zou, Caifeng; Liu, Jianqi: Security in the Internet of Things: A Review. In: 2012 International Conference on Computer Science and Electronics Engineering (ICCSEE). volume 3. IEEE, pp. 648–651, March 2012.

[Sv16] Svenda, Petr; Nemec, Matus; Seka, Peter; Kvasnovsk, Rudolf; Formane, David; Komarek, David; Matyas, Vashek: The Million-Key Question — Investigating the Origins of RSA Public Keys. In: 25th USENIX Security Symposium. USENIX, 2016.

[TF16] Tschofenig, H.; Fossati, T.: , IETF RFC 7925 Transport Layer Security (TLS) / Datagram Transport Layer Security (DTLS) Profiles for the Internet of Things. https://tools.ietf.org/html/rfc7925, 2016. (Accessed: 2019/06/17).

[UC17] US-CERT: , Alert (TA16-288A): Heightened DDoS Threat Posed by Mirai and Other Botnets. https://www.us-cert.gov/ncas/alerts/TA16-288A, 2017. (Accessed: 2019/06/17).

[VA14] Vidalis, Stilianos; Angelopoulou, Olga: Assessing Identity Theft in the Internet of Things. journal of IT Governance Practice, Vol. 2 (1): 15- 21, 2014.

[VP16] Vanhoef, Mathy; Piessens, Frank: Predicting, Decrypting, and Abusing WPA2/802.11 Group Keys. In: 25th USENIX Security Symposium. USENIX, 2016.

[VP17] Vanhoef, Mathy; Piessens, Frank: Key Reinstallation Attacks: Forcing Nonce Reuse in WPA2. In: 24th ACM Conference on Computer and Communications Security. 2017.

[Wa16] Waldvogel, Marcel: , DDoS: What we can do to prevent it. https://netfuture.ch/2016/09/ddos-what-we-can-do-to-prevent-it/, 2016. (Accessed: 2019/06/17).

[Wi14] Wi-Fi Alliance: , Wi-Fi Certified Wi-Fi Protected Setup. https://www.wi-fi.org/discover-wi-fi/wi-fi-protected-setup, 2014. (Accessed: 2019/06/17).

[Zi15] Zillner, Tobias: , ZigBee exploited: The good, the bad and the ugly. https://www.blackhat.com/docs/us-15/materials/us-15-Zillner-ZigBee-Exploited-The-Good-The-Bad-And-The-Ugly-wp.pdf, 2015. (Accessed: 2019/06/17).

Extended Abstracts

Cross-Layer Pacing for Predictably Low Age of Information

Presentation of work originally published in the Proceedings of the 2019 Workshop on Ultra-Low Latency in Wireless Networks

Andreas Schmidt[1] Stefan Reif[2] Pablo Gil Pereira[1] Timo Hönig[2] Thorsten Herfet[1]
Wolfgang Schröder-Preikschat[2]

Abstract: For dynamic systems, it is mandatory that all components operate on predictably "fresh" data. This requirement constitutes a challenge, in particular, when Internet-based or wireless communication is involved. We propose X-Pace, an approach based on cross-layer pacing, to achieve predictably low communication latency in the presence of varying network channel properties and node performance.

Keywords: Cross-Layer Optimization; Pacing; Age of Information; Low Latency; Transport Protocols

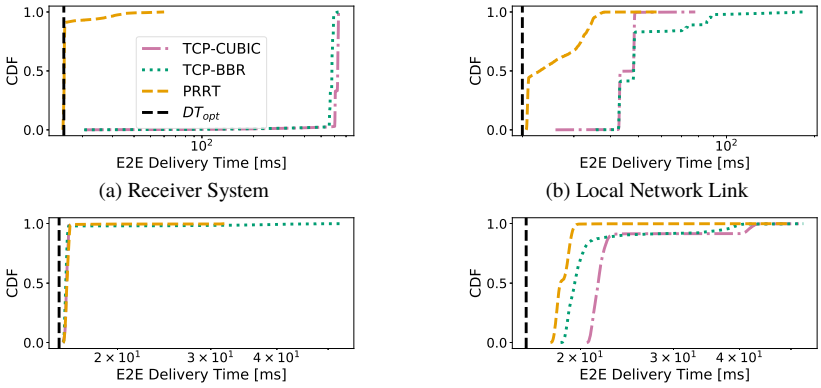
Emerging application domains such as the Internet of Things (IoT), smart factories, and inter-connected cars have initiated a transition from deeply embedded cyber-physical systems to collaborating systems. Such networked cyber-physical systems no longer operate in isolation, but use shared communication media to cooperate.

One of the key challenges is to ensure that all system components operate on "fresh" data. The age of information depends on data dependencies, scheduling, function execution times, queueing delays, and network transmission times. In modern systems, information processing chains are so long and complex that various sources of non-determinism accumulate, and potentially cause system components to operate on outdated data. The main reason for non-determinism is the dynamic behavior of networked cyber-physical systems. Besides varying network channel conditions in the Internet or wireless links, node performance depends on available resources, for example due to power constraints in embedded nodes, or contention on edge-located computing nodes.

To achieve predictably low age of information, we propose cross-layer pacing. It eliminates queueing delays, which are a main source of latency and jitter in large-scale networked systems, by making sure that (a) all processing steps run at the same speed and (b) data is produced exactly at the moment it can be consumed by the following processsing step. We have implemented this approach in a prototypical run-time system, X-Pace[3] [Sc19]. One fundamental building block is BBR [Ca16], a congestion control algorithm that utilizes the available network data rate efficiently, while keeping network-related queues empty.

[1] Saarland Informatics Campus, {andreas.schmidt,gilpereira,herfet}@cs.uni-saarland.de
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg, {reif,thoenig,wosch}@cs.fau.de
[3] X-Pace = short for cross-Layer pacing

doi:10.18420/inf2019_22



(a) Receiver System (b) Local Network Link

(c) Sender System (d) Internet Link

Fig. 1: Application-level packet delivery times depending location of the bottleneck (at system level for (a) and (c), at the network level for (b) and (d))

It simultaneously measures relevant channel properties needed for cross-layer pacing. In particular, these measurements allow our system to detect the current bottleneck component, to communicate the bottleneck pace through the entire system, and to enforce the correct speed at all layers in the system. Importantly, the pacing scheme includes the application layer. If an application constitutes the bottleneck, we detect this scenario by measuring its run-time. Otherwise, if the application is capable of running faster than appropriate, it is forced to slow down to enforce that all system components operate with the right timing. In summary, this cross-layer approach ensures that all buffers, network-based as well as node-based, remain empty.

We have implemented X-Pace in PRRT, a predictably reliable real-time transport protocol. It provides an ordered packet stream with partial reliability. The latter is based on expiry dates that are assigned to packets, incorporating the application's needs. PRRT also utilizes a hybrid error control scheme to achieve predictable reliability.

We compare X-Pace, implemented in PRRT, against TCP with the *CUBIC* and *BBR* congestion control algorithms. The results, visualized in Figure 1, demonstrate that X-Pace dynamically detects the bottleneck component and adapts appropriately. In consequence, PRRT achieves lower latency and jitter than both TCP variants.

Bibliography

[Ca16] Cardwell, Neal; Cheng, Yuchung; Gunn, C Stephen; Yeganeh, Soheil Hassas; Jacobson, Van: BBR: Congestion-based congestion control. ACM Queue, 14(5):50:20–50:53, Dec. 2016.

[Sc19] Schmidt, Andreas; Reif, Stefan; Gil Pereira, Pablo; Hönig, Timo; Herfet, Thorsten; Schröder-Preikschat, Wolfgang: Cross-Layer Pacing for Predictably Low Latency. In: Proc. 6th Intl. Worksh. on Ultra-Low Latency in Wireless Networks (Infocom ULLWN). IEEE, Apr. 2019.

Track 3 – Data Science

Data Science

Ingo Scholtes[1] Markus Strohmaier[2]

Innovations in data analytics and machine learning are key drivers for the digitalization of virtually all aspects of our everyday life. They help us to extract knowledge from large amounts of structured and unstructured data, support (automated) decision-making, generate novel business models in industry, foreshadow personalized health technologies, and transform scientific research fields thanks to the availability of new types of data and methods. T he widespread adoption of data science creates exciting challenges across all areas of computer science – including artificial intelligence, database technologies, distributed systems, computer architecture, software engineering, algorithm design, and theory -– but also raises novel societal challenges like privacy, fairness, transparency and accountability.

The Data Science track aimed to give an overview of all aspects related to the modelling, analysis, knowledge extraction, and learning from big data, with a special emphasis on recent works from the German, Austrian, and Swiss research community. The topics of interest covered data science applications in academia and industry that cross disciplinary borders, as well as works that advance the methodological foundation of data analytics and visualisation, artificial intelligence, statistical learning, and big data processing. Contributions could be submitted in either of the following categories:

- **Regular articles** were expected to present novel insights and reliable results in one of the track's topic areas and must not have been submitted or published elsewhere.
- **Extended abstracts** were expected to summarize works that have recently been published in a leading international conference or journal in the area of data science. Accepted contributions in this category will be presented during a special session "Best of Data Science made in D/A/CH".

The track received a total of 45 submissions, of which 33 fall into the extended abstract and 12 fall into the regular article category. After a thorough peer review process, we accepted 25 contributions, of which 21 are extended abstracts and four are regular articles. Hence,

[1] Chair for Data Analytics at University of Wuppertal and Head of the Data Analytics Group at the Department of Informatics at University of Zurich, scholtes@ifi.uzh.ch
[2] Chair for Methods and Theories of Computational Social Sciences and Humanities at RWTH Aachen and Scientific Coordinator for Digital Behavioral Data at GESIS - Leibniz Institute for the Social Sciences, Cologne, markus.strohmaier@humtec.rwth-aachen.de

doi:10.18420/inf2019_23

the regular article acceptance rate was approx. 33 % while the extended abstract acceptance rate was approx. 63 %.

Given the exceptionally high quality of submissions on already published works, the relatively high acceptance rate among extended abstracts was to be expected. The extended abstracts eventually included in the proceedings describe works that published in some of the world's most prestigious data science outlets, including venues like NeurIPS, SIGKDD, IEEE BigData, WWW, IEEE Big Data, Hypertext, ICDM, ICLR, SIGSPATIAL, IEEE DSAA, Machine Learning, and ESWC. We see this as testimony to the exceptional quality of research done within the German, Swiss, and Austrian data science community and we thank all authors for their valuable contributions to this track.

We further want to express our gratitude to the members of the program committee, who faced the difficult task of making a selection from a list of high-quality submissions. The program comittee consisted of the following 22 members:

- Christian Bauckhage, Fraunhofer-Institut für Intelligente Analyse- und Informationssysteme IAIS
- Martin Becker, Universität Würzburg
- Christian Bizer, Universität Mannheim
- Rebekka Burkholz, Harvard University
- Rainer Gemulla, Universität Mannheim
- Michael Granitzer, Universität Passau
- Stephan Günnemann, Technische Universität München
- Denis Helic, Technische Universität Graz
- Andreas Hotho, Universität Würzburg
- Eyke Hüllermeier, Universität Paderborn
- Enkelejda Kasneci, Eberhard-Karls-Universität Tübingen
- Elisabeth Lex, Technische Universität Graz
- Alexander Munteanu, Technische Universität Dortmund
- Jürgen Pfeffer, Technische Universität München
- Matthias Rottmann, Bergische Universität Wuppertal
- Markus Schedl, Johannes Kepler Universität Linz
- Frank Schweitzer, ETH Zürich
- Bernhard Seeger, Philipps-Universität Marburg
- Martin Theobald, Université du Luxembourg
- Claudia Wagner, GESIS - Leibniz Institut für Sozialwissenschaften
- Robert West, École Polytechnique Fédérale de Lausanne (EPFL)
- Katharina Anna Zweig, Techniche Universität Kaiserslautern

Full Papers

EClaiRE: Context Matters! – Comparing Word Embeddings for Relation Classification

Lena Hettinger[1] Albin Zehe[1] Alexander Dallmann[1] Andreas Hotho[1]

Abstract: In recent years, there has been an increasing interest in the task of relation classification, which aims to label a relation between two semantic entities. In this work, we investigate how domain-specific information influences the performance of ClaiRE, an SVM-based system combining manually crafted features with word embeddings. To this end, we experiment with a wide range of word embeddings and evaluate on one general and two scientific relation classification datasets. We release all of our code for relation classification and data for scientific word embeddings to enable the reproduction of our experiments.[2]

Keywords: word embedding; relation classification; context sensitive; domain specific

1 Introduction

Finding an appropriate representation for a word is a challenging task in Natural Language Processing, especially considering the fact that words can have multiple meanings. Word ambiguity becomes most apparent when looking at datasets from different domains. For example, the word *string* can either denote a "string of cotton" or a "string of characters", depending on whether it appears in its most common or a domain specific context like computer science. In this paper, we want to examine the role of ambiguous or domain-specific expressions for the relation classification task (cf. Sect. 4).

The goal of relation classification is to label the semantic relation between two selected entities in a sentence. There are two reasons why this is a fitting task to examine word representations. First, there exist indications that a semantically correct representation is important for good performance [He18]. Second, it is obvious that a certain relation described between two entities might not correspond to the most general meaning of a word, as exemplified in Fig. 1. Identifying words only with their most common meaning, the sentence in the example would not make sense and the expressed relation would be unclear.

Since our focus is more on understanding the influence of domain-specific word senses than on providing a new state of the art, we decided to use an SVM-based model rather than a neural network, which requires far less computational power and therefore enables us to

[1] University of Wuerzburg, DMIR Group, Am Hubland, 97074 Würzburg, Germany {hettinger,zehe,dallmann, hotho}@informatik.uni-wuerzburg.de

[2] https://gitlab2.informatik.uni-wuerzburg.de/dmir/eclaire

doi:10.18420/inf2019_24

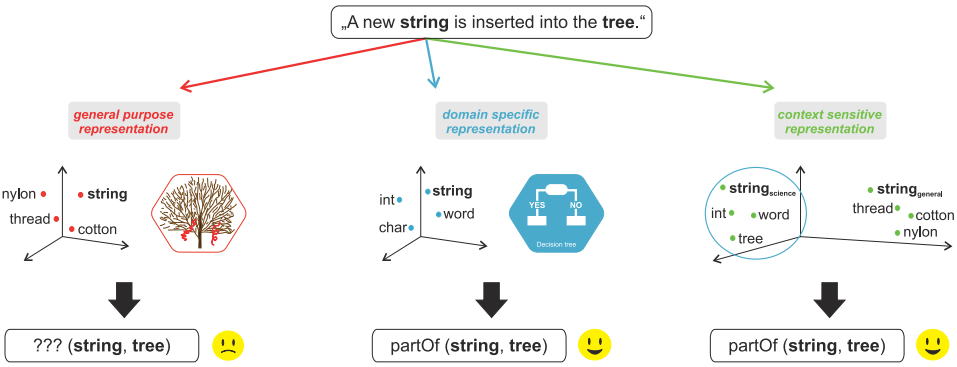


Fig. 1: Problems in relation classification arising from word representations that do not reflect domain-specific meaning.

investigate a wider range of settings. Specifically, we use ClaiRE [He18], an SVM based on both hand-crafted features and word embeddings. ClaiRE has already been shown to be highly dependent on the word embeddings used to encode the input, making it a suitable model for our research.

We investigate a wide range of word embeddings, which can generally be partitioned into three classes: (a) Publicly available static embeddings trained on general corpora, (b) our own domain specific, static embedding trained on a corpus of scientific articles and (c) publicly available context-sensitive embeddings trained on a general corpus.

We show that domain-specific word embeddings outperform general ones on the scientific domain, and, conversely, that specialised word embeddings can not be transferred to the general domain. We also find that context-sensitive embeddings outperform both types of static embeddings on each domain, but further improvements can be reached by combining them with domain-specific embeddings.

2 Related Work

Relation classification is an interesting topic of research, as relations between pairs of entities are studied across a wide range of domains. It was the topic of several challenges, e. g. SemEval-2010 (general domain) [He09] and SemEval-2018 (scientific domain) [Gá18]. Although neural network based approaches currently claim state of the art on both domains [RHZ18, Wa16], SVM based approaches with lexical and semantic features have shown competitive performance in the past [He18, RH10]. For a more in-depth coverage of related work on relation classification we refer to [He18, RHZ18]. In this work we investigate the dependency of relation classification on feature and task domain. Hence, we do not focus on achieving state-of-the-art results.

Word embeddings have been proven effective for a wide range of NLP tasks [Ki14, Ma14]. Consequently, much work has been done on word embeddings in recent years. We will provide an overview of some commonly used models in Sect. 3.

Although word embeddings often improve the performance of downstream tasks, word embeddings derived from general corpora can be suboptimal if used in specialised domains [Ta14]. For example, word embeddings trained on a domain-specific corpus improve relation classification performance in the scientific domain [He18]. However, large corpora needed for training a word embedding model from scratch are not always available. As a result, some work has been done on leveraging multiple corpora by training cross-domain embeddings [BMiK15, YLZ17]. Another direction of research focuses on adapting pre-trained language models to a specific domain [HR18, LL13, Ni17, Yu17].

Commonly used word embeddings provide a single static vector for every word in the vocabulary. As a result, different word senses are not accurately represented. This issue has been addressed by modifying existing models to represent a word by multiple sense vectors [IPN15, CLS14, JP15]. However, these models still suffer from limitations, for example by relying on a limited semantic network [JP15]. A recent line of research addresses these limitations by learning context-sensitive embeddings that compute a word representation dependent on the context, e.g. the sentence [CCP18]. Evaluation of context-sensitive embeddings is mostly focused on the general domain [ABV18, Mc17, Pe18]. In contrast we aim to investigate the suitability of general context-sensitive embeddings for a domain specific task and compare the performance to domain-specific static embeddings.

3 Background: Word Embeddings

In order to analyse the connection between the domain of relation classification and word embeddings, we first need to take a look at the data and models used to create them. This will later enable us to pose some hypotheses as to why certain embeddings work better for one or the other context. This chapter provides an overview over the different word embeddings we use in this work. We classify them into three groups: Traditional embeddings, domain-specific embeddings and context-sensitive embeddings, as shown in Tab. 1.

3.1 Traditional Embeddings

The first group of embeddings are publicly available sets of vectors that are commonly used in NLP. There exist multiple algorithms for the creation of these embeddings, where some of the most commonly used are word2vec [Mi13], GloVe [PSM14] and FastText [Bo17]. We use publicly available vectors trained with each of these algorithms, as well as ConceptNet Numberbatch [SCH17], which is a combination of the three aforementioned embeddings and the knowledge graph ConceptNet. The first part of Tab. 1 provides some details about the

traditional embeddings used in this work. We retrieve all of these through `gensim-data`[3], a data repository for pre-trained NLP models.

WE	dim	Origin	size
w2v	300	Google	100.0
GloVe	300	WP, Gigaword	6.0
FastText	300	WP, Gigaword, ...	16.0
CNB	300	w2v, GloVe, ...	n.a.
$w2v_{arXiv}$	300	arXiv	0.7
ELMo	3072	WP, WMT	5.5
Flair	4096	WMT	0.8

Tab. 1: Details about the pretrained word embeddings used in this paper. WP: Wikipedia, WMT: Workshop on Statistical Machine Translation. Size of data sets is given in billion tokens.

3.2 Domain-specific Embeddings

To generate domain-specific $w2v_{arXiv}$ embeddings, we use word2vec on a large corpus of scientific papers. We downloaded LaTeX sources for all papers published in 2016 on arXiv.org using the provided dumps.[4] We converted LaTeX sources to plain text using a manually crafted set of regular expressions. We refer to our source code for details about the conversion and to [He18] for further details about constructing scientific embeddings.

3.3 Context-Sensitive Embeddings

Context-sensitive embeddings have been on the rise since the publication of CoVe [Mc17]. They represent a change of paradigm, as words are not described by static vectors but are assigned a different vector depending on the sentence they appear in. The advantage of this approach is that ambiguous words do not have to be resolved separately. Instead the model is able to distinguish between the different meanings using the word's context. In this work we compare traditional embeddings with the two currently best-performing context-sensitive models: ELMo, which has shown to perform better than CoVe, and Flair, which outperforms ELMo for some tasks in conjunction with character features and/or traditional embeddings. These embeddings make up the third part of Tab. 1.

[3] `https://github.com/RaRe-Technologies/gensim-data`
[4] `https://arxiv.org/help/bulk_data`

3.3.1 ELMo

ELMo (Embeddings from Language Models) [Pe18] is a deep bidirectional language model that produces character based word vectors from its internal states. More specifically, ELMo computes multiple representations with different amounts of context and semantics in its layers.

First, the model builds context insensitive word representations ($ELMo_0$) by applying a character based CNN on every word in the sentence. The context sensitive embeddings ($ELMo_1$, and $ELMo_2$ respectively) stem from a 2-layer biLSTM [HS97, SP97] that takes the previously computed word representations ($ELMo_0$) as input. Ultimately, the biLM provides three layers of representations for any input token, forming a hypercolumn of the three vectors ELMo = [$ELMo_0$, $ELMo_1$, $ELMo_2$].

ELMo claims that lower-level LSTM states capture aspects of syntax, while higher-level states model aspects of word meaning in context. It has proven its success on different tasks such as question answering, semantic role labeling and named entity extraction. However, it has not been applied to the task of relation classification before. To the best of our knowledge, the performance of ELMo across different domains of data has not been researched yet explicitly. For our experiments, we use an officially published ELMo model[5] that has been pre-trained on a 5.5 billion token general dataset and produces context sensitive token representations ELMo $\in \mathbb{R}^{3072}$.

3.3.2 Flair

Flair [ABV18] is a context-sensitive model which claims to outperform ELMo on tasks such as named entity recognition and chunking. Similar to ELMo it leverages the internal states of a trained language model, but uses characters as atomic units for a 1-layer biLSTM. Flair is thus trained without any explicit notion of words and at each point in the sequence predicts the next character. This is different from the character-aware LM used by ELMo, which operates on word level and character convolutions.

In this work we utilise Flair embeddings trained on the 1-billion word corpus [Ch13].[6] A Flair vector consists of the 2048 hidden states of a forward and backward LSTM, which can be described as a hypercolumn: Flair = [$Flair_f$, $Flair_b$] $\in \mathbb{R}^{4096}$. We try different combinations of Flair vectors (as well as for ELMo), to see which best fits the relation classification task in Sect. 6.3.

[5] `https://allennlp.org/elmo`, (Original 5.5B)

[6] `https://github.com/zalandoresearch/flair/blob/master/resources/docs/TUTORIAL_WORD_EMBEDDING.md`, (news-forward/ -backward)

4 Task: Relation Classification

We will now describe the task of relation classification and the model we utilise to investigate the effect of task and feature domain in detail.

4.1 Task Description

The goal of relation classification is to classify semantic relations between entities into a predefined set of categories. In order to further illustrate the problem we are dealing with in this paper, we picture two specific relation samples for each domain, general and scientific text, in Tab. 2. Relations are marked as reversed, if the order their entities appear in does not match the class order (cf. Sect. 4.2).

Across domains, some tokens have ambiguous meaning. For example, while the word "paper" is commonly associated with a material, in the scientific domain it will mostly stand for a publication. As words may be linked to specific relation classes, representing different appearances with a single static vector and thus ignoring the context they appear in might lead to misclassification.

Domain	Label	Sample
General	Component-Whole	The tailpiece anchors the strings to the lower **bout** of the **violin** by means of the tailgut.
General	Instrument-Agency (rev)	Stanford **researchers** have coated **paper** with carbon nanotubes [. . .].
Science	Result	**Combination methods** are an effective way of improving **system performance**.
Science	Usage (rev)	In this paper we describe a speaker dependent system for predicting segmental **duration** from **text** [. . .].

Tab. 2: Examples for relation classification samples from the general and scientific domain. Relation entities are denoted by bold font.

4.2 Feature Extraction: ClaiRE

We will use ClaiRE[7] as a base system for relation classification. ClaiRE is based on an SVM trained on a combination of word embeddings with manually crafted features. We selected this method as the original paper has shown that both the manual features and the word embeddings are critical for the performance on SemEval-2010 Task 8. Thus, it is reasonable to keep the manual features fixed while swapping different word embeddings to compare their performance.

[7] `https://gitlab2.informatik.uni-wuerzburg.de/dmir/claire`

In Tab. 3, we provide a short overview on hand-crafted lexical features constructed from text and numeric features based on word embeddings, for more details see [He18]. When constructing features for relation classification, the relevant parts of a sentence consist of the two entities that are part of a relation and their context, meaning the words in between entities. We distinguish between a *start* and *end entity* of a relation. If the start entity appears after the end entity within a sentence, the direction of a relation is **reversed**, as can be seen in Tab. 2.

Some features are slightly modified in this work, noted by a star in Tab. 3. In contrast to [He18], we did not utilise SpaCy[8] for preprocessing text and treated *dist* and *sim* as numeric features instead of boolean to provide more information. We utilise the WordNet Lemmatizer of nltk[9] for lemmatisation and the Stanford POS Tagger [To03] for Part-of-Speech (POS) tags. Before computing features, all lemmatised context words below a certain threshold were discarded to limit the vocabulary. Preliminary tests have shown that the optimal lemma frequency is 5 for both datasets.

Feature Set	Description
*bow**	BOW (lemmatised) from context
*pos**	Stanford POS tags from context
*pospath**	concatenated POS tags from context
*dist**	number of words in context
lc	Levin classes of verbs in context
ents	entity (head) without order
startEnt	entity (head)of relation start
endEnt	entity (head) of relation end
c	embedding vector of context
e_1	embedding vector of first entity
e_2	embedding vector of second entity
*sim**	similarity score of two entity vectors
simb	similarity bucket of similarity score

Tab. 3: Generated features for use in relation classification, grouped by type: lexcial context, lexical entity and embedding features. Features which differ slightly from [He18] are marked by *.

5 Datasets: General and Scientific Relations

Since we want to compare the performance of different word embeddings across domains, we need datasets from different domains of text. There have been multiple SemEval tasks concerned with the task of relation classification in the past, some on general corpora and some on data from specialised domains. We use the dataset from SemEval-2010 Task 8 (SE10-8) as an example of a "general" corpus and the dataset from SemEval-2018 Task 7 for a specialised corpus, in particular for the scientific domain.

[8] `https://spacy.io/`
[9] nltk.stem.wordnet.WordNetLemmatizer

SE10-8 consists of 10 717 samples of semantic relations between nominals in a sentence, collected by pattern-based web search, where 8000 samples are used for training (SE10-8_{train}) and 2717 as a test set (SE10-8_{test}). Each sample is labelled with one of 10 classes, see [He09] for a detailed description. In addition to the relation label, the direction of a relation has to be predicted for this task. We decided to model the direction as part of the label (e. g. *Cause-Effect-Reverse*) instead of using a two-stage approach (i.e., predicting the label and the direction separately) as initial experiments have shown that this performs better for ClaiRE.

The relation classification task (task 7) of SemEval-2018[10] is comprised of two subtasks. In the first subtask participants were provided with 1228 training samples and a test set (SE18-7_{clean}) with 355 relations, where both entities and relations were manually labelled. The second subtask consists of 1245 training samples and a different test set (SE18-7_{noisy}) with 355 relations, but here entities have been extracted automatically, thus introducing noise. Samples for both subtasks stem from abstracts from the ACL Anthology Corpus.

Combining both training sets has been shown to improve performance on both subtasks [He18], thus we form our final training set (SE18-7_{train}) by combining the training samples from both tasks. The final training set then has 2473 samples and each sample belongs to one of the six domain-specific classes in Tab. 4. As the relation direction is not part of the classification task it can be utilised as a feature in this case.

Note that both test sets SE18-7_{clean} and SE18-7_{noisy} contain classes (TOPIC, COMPARE) that are heavily underrepresented. This leads to some artifacts in the evaluation score for this dataset, which we will discuss in Sect. 6.2.

label	**SE18-7_{clean}**		**SE18-7_{noisy}**	
COMPARE	21	5.9 %	3	0.8 %
MODEL-F.	66	18.6 %	75	21.1 %
PART_W.	70	19.7 %	56	15.8 %
RESULT	20	5.6 %	29	8.2 %
TOPIC	3	0.8 %	69	19.4 %
USAGE	175	49.3 %	123	34.6 %

Tab. 4: Distribution of class labels for the SE18-7 datasets with absolute values and relative frequency.

6 Comparison of Embeddings and Datasets from Different Domains

We will now describe the experimental setup used in our relation classification evaluation and the results we obtained.

[10] https://competitions.codalab.org/competitions/17422

6.1 Experimental Setup

In order to assess the relative quality of different embeddings for a domain, we follow the setting in [He18], using an rbf-SVM as a base classifier and the features described therein with changes as noted in Sect. 4.2. We keep the hand-crafted features fixed while varying the embedding-based features, constructing them from our different embeddings. We also experiment with using a combination of multiple embeddings. In this case, we construct all embedding features using the concatenation of the embeddings.

To further strengthen our model, we use an ensemble of 10 SVMs with shuffled training data. As we utilise the probability estimates of an SVM to predict test labels [WLW04], we average over all probabilities in the ensemble and predict the class with the highest score. We use macro-averaged F1-score to evaluate our models, as the rating in both of the SemEval tasks was based on this score. We made use of the respective official evaluation scripts to compute the scores.

6.2 Classification Results

We report overall results for all three datasets in Tab. 5 before taking a closer look at different embeddings.

The first line reports the best result achieved on the respective datasets by any previously presented system. The best systems rely on rather complicated neural networks that are specifically tuned to the task, requiring large amounts of computational power. To enable fair comparison against an SVM-based system, the next line shows the best SVM so far as a baseline. Taking only lexical features into consideration and excluding word embeddings completely (w/o WE), ClaiRE exhibits insufficient performance, once again proving the worth of word embeddings for the task of relation classification. On the other hand, using only word embedding features and ignoring the lexical features (only WE) already performs rather well.

The next part of the table shows the performance of ClaiRE with static word embeddings (word2vec, GloVe, CNB and FastText) in combination with hand-crafted features. The embeddings have been pre-trained on large general corpora (see Sect. 3). Our results show that by utilising these embeddings, ClaiRE performs quite well on the SE10-8 dataset from the general domain, but the performance deteriorates on the scientific datasets from SE18-7.

The opposite effect can be found for the domain-specific embedding. As expected, $w2v_{arXiv}$ greatly outperforms traditional WEs on the scientific domain. On the general SE10-8 data, however, the scientific embedding performs far worse than the general embeddings.

The final part of the table shows the performance of the context-sensitive embeddings ELMo and Flair. For this summary we utilise the first context-sensitive ELMo-layer ($ELMo_1$) and

the Flair backward-layer (Flair$_b$). We will take a closer look at embedding-layers in Sect. 6.3 and show the reason for that decision. While both ELMo and Flair consistently perform well on both domains, ELMo achieves better scores on both the SE10-8 and SE18-7$_{\text{clean}}$ dataset.

Model	**SE10-8**	**SE18-7** clean	**SE18-7** noisy
best	*88.00*[a]	*81.72*[b]	*90.40*[b]
best SVM	82.19[c]	75.11[d]	81.44[d]
w/o WE	73.15	68.58	74.10
only WE	82.42[e]	76.27[e]	**85.79**[f]
w2v	78.46	72.28	77.56
GloVe	77.45	67.03	81.21
CNB	79.20	72.18	79.12
FastText	79.50	69.21	81.57
w2v$_{\text{arXiv}}$	74.90	77.76	84.47
ELMo$_1$	83.22	80.34	84.21
Flair$_b$	79.01	77.32	85.38
best 2-WE	**83.81**[g]	**81.13**[h]	85.63[g]

[a] [Wa16] [b] [RHZ18] [c] [RH10] [d] [He18]
[e] ELMo$_1$ [f] w2v$_{\text{arXiv}}$ [g] ELMo$_1$/Flair$_b$
[h] ELMo$_1$/w2v$_{\text{arXiv}}$

Tab. 5: Results from relation classification on three datasets using different embeddings. Results are given as macro-averaged F1-scores. The best result achieved by an SVM for each dataset is marked in bold. Footnotes denote the embedding used on the respective datasets.

We also evaluated combinations of lexical features and two embeddings. The results are shown in the last line of the table (best 2-WE)[11]. Again, context-sensitive ELMo embeddings contribute substantially to a very good performance, appearing in the best WE-pair of each dataset.

Overall, ELMo is the best-performing embedding for two out of three datasets. The automatically built dataset SE18-7$_{\text{noisy}}$ forms an exception, as domain-specific w2v$_{\text{arXiv}}$ vectors perform best for this task.But as mentioned in Sect. 5, the label distributions of the scientific test sets are heavily skewed, thus leaving macro-F1 vulnerable to performance shifts on very small classes. We therefore investigated micro-averaged F1-score, as it aggregates the contributions of all classes, without noting class imbalance. In our setting, micro-F1 consistently scores approximately 1% above macro-F1, emphasizing the good results of our models on big classes. The only special case is, as mentioned above, w2v$_{\text{arXiv}}$ on SE18-7$_{\text{noisy}}$, with a macro-F1 of 85.79 and micro-F1 of only 83.94. By contrast, ELMo$_1$ delivers results of 85.63 macro- and 86.20 micro-F1; in other words, similar macro-F1 but

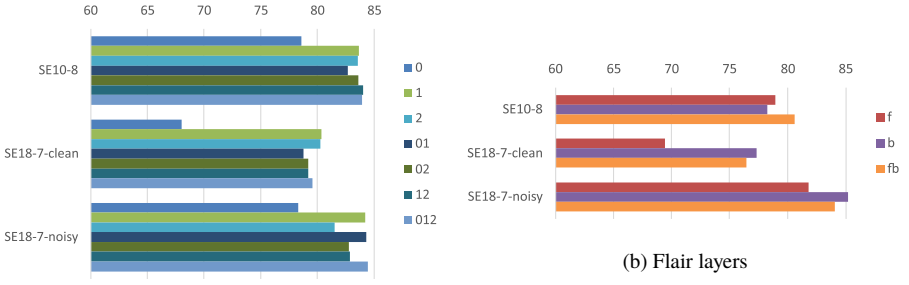
[11] We evaluated combinations of three embeddings as well, but results did not improve.

quite different micro-F1. Hence, the result of $w2v_{arXiv}$ on $SE18\text{-}7_{noisy}$ must be a case of overfitting on small classes and overestimating classifier performance by usage of macro-F1. Note that we still outperform the best previous SVM for all data sets if we add contextual embeddings to ClaiRE.

6.3 Analysis of Contextual Layers

After comparing different word embeddings and embedding types in the previous section, we will now look at the best configuration of context-sensitive embeddings. Both utilised context-sensitive models produce word vectors from the internal states of different LM-layers. As the best combination of ELMo-layers depends on the task at hand [Pe18], we investigate relation classification performance for different configurations of ELMo- and Flair-layers.

As shown in Fig. 2a, the static ELMo-layer $ELMo_0$ performs notably worse for all three datasets, while $ELMo_1$ is the best singular layer. Combining different layers does not change performance for relation classification considerably. For Flair we find that there exists no clear advantage for a layer combination across datasets (cf. Fig. 2b). As the backwards-layer $Flair_b$ alone scores best for two out of three tasks, we chose it for our experiments in Sect. 6.2.



(a) ELMo layers

(b) Flair layers

Fig. 2: Results as macro-F1 for different combinations of layers from context-sensitive embeddings (including lexical features).

6.4 Analysis of Nearest Neighbours

To illustrate the behaviour of different embedding types on two domains, general and scientific, we present nearest neighbours in the embedding space for some ambiguous words in Tab. 6. We determine closeness by means of cosine similarity and report five nearest neighbours of any vocabulary entry in the case of w2v and $w2v_{arXiv}$. We additionally compute the closest word in the ELMo embedding space for a) a token appearance in SE10-8 and b) an appearance in SE18-7 and report their associated sentences.

As shown in Tab. 6, ambiguous word have a different meaning in the general and the scientific domain, as is evidenced by their neighbourhood in the respective domain embeddings. In contrast, ELMo embeds words into a vector space depending on their context, enabling different nearest neighbours matching their word sense. We assume that this distinction between words senses contributes to a mapping of entities to relations.

WE	**word (in context)**	**nearest neighbours**
w2v	tree	trees, pine tree, oak tree, evergreen tree, fir tree
$w2v_{arXiv}$		trees, subtree, subtrees, leaf, graph
ELMo	a) An oak **tree** grows from an acorn.	Winter is here and the little fir **tree** stands lonely in the forest.
	b) We use decision **trees** to learn the controllers.	This paper describes novel and practical Japanese parsers that uses decision **trees**.
w2v	string	spate, slew, rash, litany, flurry
$w2v_{arXiv}$		strings, superstring, worldsheet, brane, worldsheets
ELMo	a) A **string** of pack ponies trotted through the pines behind them.	I remembered about a **string** of rosary beads [. . .]
	b) One is **string** similarity based on edit distance.	We take a selection of both bag of words and segment order sensitive **string** comparison methods [. . .]

Tab. 6: Most similar words for different static embeddings and ELMo.

7 Conclusion

Our intuition was that general word embeddings would fail to capture the meaning of some words for relation classification on scientific data, while specialised word embeddings would in turn fail to work outside their domain. This intuition is supported by our results. We also hypothesised that context-sensitive word embeddings would be able to generalise across domains, as they can model multiple meanings of a word and distinguish them by their current context. This assumption also seems to hold true, as evidenced by the consistently great performance of ELMo and the good performance of Flair.

Overall, ECLaiRE - our combination of ClaiRE and ELMo - a simple rbf-SVM with a few hand-coded features and context-sensitive word embeddings, is able to outperform the best SVM classifiers so far and even achieves similar results to a complex neural network architecture for the $SE18\text{-}7_{clean}$ task. Thus, it may be useful to introduce context-sensitive word embeddings, especially ELMo, to more relation classification datasets from different domains.

Bibliography

[ABV18] Akbik, Alan; Blythe, Duncan; Vollgraf, Roland: Contextual String Embeddings for Sequence Labeling. In: COLING. pp. 1638–1649, 2018.

[BMiK15] Bollegala, Danushka; Maehara, Takanori; ichi Kawarabayashi, Ken: Unsupervised Cross-Domain Word Representation Learning. In: ACL. The Association for Computer Linguistics, pp. 730–740, 2015.

[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.

[CCP18] Camacho-Collados, José; Pilehvar, Mohammad Taher: From Word to Sense Embeddings: A Survey on Vector Representations of Meaning. Journal of Artificial Intelligence Research, 63:743–788, 2018.

[Ch13] Chelba, Ciprian; Mikolov, Tomas; Schuster, Mike; Ge, Qi; Brants, Thorsten; Koehn, Phillipp; Robinson, Tony: , One Billion Word Benchmark for Measuring Progress in Statistical Language Modeling, 2013.

[CLS14] Chen, Xinxiong; Liu, Zhiyuan; Sun, Maosong: A unified model for word sense representation and disambiguation. In: EMNLP. pp. 1025–1035, 2014.

[Gá18] Gábor, Kata; Buscaldi, Davide; Schumann, Anne-Kathrin; QasemiZadeh, Behrang; Zargayouna, Haïfa; Charnois, Thierry: SemEval-2018 Task 7: Semantic Relation Extraction and Classification in Scientific Papers. In: SemEval@NAACL-HLT. pp. 679–688, 2018.

[He09] Hendrickx, Iris; Kim, Su Nam; Kozareva, Zornitsa; Nakov, Preslav; Ó Séaghdha, Diarmuid; Padó, Sebastian; Pennacchiotti, Marco; Romano, Lorenza; Szpakowicz, Stan: Semeval-2010 task 8: Multi-way classification of semantic relations between pairs of nominals. In: Proceedings of the Workshop on Semantic Evaluations: Recent Achievements and Future Directions. pp. 94–99, 2009.

[He18] Hettinger, Lena; Dallmann, Alexander; Zehe, Albin; Niebler, Thomas; Hotho, Andreas: ClaiRE at SemEval-2018 Task 7: Classification of Relations using Embeddings. In: Proceedings of International Workshop on Semantic Evaluation. 2018.

[HR18] Howard, Jeremy; Ruder, Sebastian: Universal Language Model Fine-tuning for Text Classification. In: ACL. Association for Computational Linguistics, 2018.

[HS97] Hochreiter, Sepp; Schmidhuber, Jürgen: Long Short-Term Memory. Neural Computation, 9(8):1735–1780, November 1997.

[IPN15] Iacobacci, Ignacio; Pilehvar, Mohammad Taher; Navigli, Roberto: Sensembed: Learning sense embeddings for word and relational similarity. In: COLING. volume 1, pp. 95–105, 2015.

[JP15] Johansson, Richard; Pina, Luis Nieto: Embedding a semantic network in a word space. In: NAACL-HLT. pp. 1428–1433, 2015.

[Ki14] Kim, Yoon: Convolutional Neural Networks for Sentence Classification. In: EMNLP. pp. 1746–1751, 2014.

[LL13] Labutov, Igor; Lipson, Hod: Re-embedding words. In: ACL. pp. 489–493, 2013.

[Ma14] Marco Baroni, Georgiana Dinu, Germán Kruszewski: Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors. COLING, 1:238–247, 2014.

[Mc17] McCann, Bryan; Bradbury, James; Xiong, Caiming; Socher, Richard: Learned in Translation: Contextualized Word Vectors. In: Advances in Neural Information Processing Systems. 2017.

[Mi13] Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Greg S; Dean, Jeff: Distributed Representations of Words and Phrases and their Compositionality. In: NIPS, pp. 3111–3119. Curran Associates, Inc., 2013.

[Ni17] Niebler, Thomas; Becker, Martin; Pölitz, Christian; Hotho, Andreas: Learning Semantic Relatedness from Human Feedback Using Relative Relatedness Learning. In: ISWC. 2017.

[Pe18] Peters, Matthew E.; Neumann, Mark; Iyyer, Mohit; Gardner, Matt; Clark, Christopher; Lee, Kenton; Zettlemoyer, Luke: Deep Contextualized Word Representations. In: NAACL-HLT. pp. 2227–2237, 2018.

[PSM14] Pennington, Jeffrey; Socher, Richard; Manning, Christopher D: Glove: Global Vectors for Word Representation. In: EMNLP. volume 14, pp. 1532–1543, 2014.

[RH10] Rink, Bryan; Harabagiu, Sanda: Utd: Classifying semantic relations by combining lexical and semantic resources. In: Proceedings of the 5th International Workshop on Semantic Evaluation. pp. 256–259, 2010.

[RHZ18] Rotsztejn, Jonathan; Hollenstein, Nora; Zhang, Ce: , ETH-DS3Lab at SemEval-2018 Task 7: Effectively Combining Recurrent and Convolutional Neural Networks for Relation Classification and Extraction, 2018.

[SCH17] Speer, Robert; Chin, Joshua; Havasi, Catherine: , ConceptNet 5.5: An Open Multilingual Graph of General Knowledge, 2017.

[SP97] Schuster, Mike; Paliwal, Kuldip K: Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing, 45(11):2673–2681, 1997.

[Ta14] Tang, Duyu; Wei, Furu; Yang, Nan; Zhou, Ming; Liu, Ting; Qin, Bing: Learning Sentiment-Specific Word Embedding for Twitter Sentiment Classification. In: COLING. pp. 1555–1565, 2014.

[To03] Toutanova, Kristina; Klein, Dan; Manning, Christopher D; Singer, Yoram: Feature-rich part-of-speech tagging with a cyclic dependency network. In: HLT-NAACL. pp. 173–180, 2003.

[Wa16] Wang, Linlin; Cao, Zhu; de Melo, Gerard; Liu, Zhiyuan: Relation Classification via Multi-Level Attention CNNs. In: COLING. volume 1, pp. 1298–1307, 2016.

[WLW04] Wu, Ting-Fan; Lin, Chih-Jen; Weng, Ruby C.: Probability Estimates for Multi-class Classification by Pairwise Coupling. Journal of Machine Learning Research, 5(Aug):975–1005, 2004.

[YLZ17] Yang, Wei; Lu, Wei; Zheng, Vincent: A Simple Regularization-based Algorithm for Learning Cross-Domain Word Embeddings. In: EMNLP. pp. 2898–2904, 2017.

[Yu17] Yu, Liang-Chih; Wang, Jin; Lai, K. Robert; Zhang, Xue-Jie: Refining Word Embeddings for Sentiment Analysis. In: EMNLP. pp. 534–539, 2017.

Do We Need Real Data? – Testing and Training Algorithms with Artificial Geolocation Data

Jan Kaiser[1] Kai Bavendiek[2] Sibylle Schupp[3]

Abstract: As big data becomes increasingly important, so do algorithms that operate on geolocation data. Privacy requirements and the cost of collecting large sets of geolocation data, however, make it difficult to test those algorithms with real data. Artificially generated data sets therefore present an appealing alternative. This paper explores the use of two types of neural networks as generators of geolocation data and introduces a method based on the Turing Test to determine whether generated geolocation data is indistinguishable from real data. In an extensive evaluation we apply the method to data generated by our own implementation of neural networks as well as the widely used BerlinMOD generator on the one hand, the four most prominent data sets of real geolocation data covering at total of 65 million records on the other hand. The experiments show that in eleven of twelve cases artificial data sets can be told from real ones. We conclude that, at present, the generators we tested provide no safe replacement for real data.

Keywords: geolocation data; artificial data; data generation; neural networks generators; data quality

1 Introduction

With increasing interest in *Big Data* in recent years, interest in geolocation data – collections of coordinates, such as latitudes and longitudes – has increased as well. Geolocation data may find use in areas, such as traffic planning, recommender systems, market research, map creation, location privacy, and autonomous driving. As do algorithms that are capable of analysing geolocation data and drawing meaningful conclusions from it. In order to develop and test such algorithms, developers and researchers, alike, need access to geolocation data sets. Unfortunately, only a very limited amount of real geolocation data sets is freely available because of privacy requirements and other difficulties associated with their collection. Real geolocation data sets are also not very flexible, as their sizes, covered areas, logged entities, etc. cannot easily be changed to meet anyone's needs.

Artificially generated data sets promise to solve the above problems. In theory, it is easily possible to generate data sets off various sizes, for various locations, with various types of entities, all of which could be adjusted to produce exactly the kind of data set a given

[1] Hamburg University of Technology; Hamburg, Germany; jan.kaiser@tuhh.de
[2] Hamburg University of Technology; Hamburg, Germany; kai.bavendiek@tuhh.de
[3] Hamburg University of Technology; Hamburg, Germany; schupp@tuhh.de

doi:10.18420/inf2019_25

application requires. However, the question remains how to generate realistic geolocation data. The other question is how to determine whether generated data is realistic. We propose an approach for determining the realism of geolocation data based on the Turing Test. We use this approach to test the BerlinMOD generator and two Artificial Neural Network generators on the realism of the data they produce.

This paper gives a brief outline of existing work in the field in Section 2. In Section 3, we present two novel methods of generating geolocation data using Artificial Neural Networks. We evaluate[4] both methods and the commonly used BerlinMOD generator using our approach to assessing the quality of generated data based on the Turing Test in Section 4. Lastly in Section 5, we conclude that the presented generators are not fit for use in research. We also conclude that our Turing-based approach for determining the realism of data delivers good results. We finish by making proposals for future work.

2 Related Work

From the previous decade, three algorithmic generators of geolocation data – BerlinMOD [DBG09], Brinkhoff's generator [Br03], and SUMO [Kr12] – have been widely accepted as sources of data for testing algorithms. SUMO and Brinkhoff's generator produce what research refers to as short-term data, i.e. observations of otherwise anonymous entities while they move from one place to another. BerlinMOD produces so-called long-term data in which entities are observed for a longer timeframe, regardless of whether they are currently moving between places or remain stationary. We are not aware of any further developments in the field of geolocation data generators in recent years.

As other means of generating data, Artificial Neural Networks have recently gained in popularity. In the area of sequenced data Recurrent Neural Networks have demonstrated impressive results producing authentic sequences, for example of text [Go17, Te17]. The text generation approach by Gorner [Go17] serves as the basis for our approach to generating sequences of geolocation data. A different type of Neural Network, Generative Adversarial Networks, have been shown to work well for generating realistic images [RMC15, Zh17]. In future work a similar approach could be used to plot realistic geolocation data on an image of a map. Both approaches have also been crossed, for example in the implementation of *C-RNN-GAN* [Mo16], which employs an architecture of recurrent Generative Adversarial Networks to produce realistic music. A similar approach based on recurrent Generative Adversarial Network is pursued in this paper.

Several metrics for comparing data sets and their quality exist, for example in the area of location privacy [MBT11]. However we are not aware of any works that devise means of evaluating the quality of artificial data in a general scenario.

[4] Implementations available at `https://github.com/LordHelmchen324/real-vs-synthetic-geospatial`

While our work explores the possibilities of generating raw data that is indistinguishable from real recorded data, there has been other work investigating formal models to describe real-world patterns found in geolocation data [Zh16]. Rhee et al., for example, have found that the movement of people in a city can be modelled mathematically as Levy walks [Rh11]. Models like this hold great potential for understanding the behaviour of people in a city, and they could also find use in generators of geolocation data.

3 Generators

Two different kinds of generators are distinguished for this paper: algorithmic generators and Artificial Neural Networks. Algorithmic generators use assumptions about the real-world behaviour of entities to produce realistic spatio-temporal data. The assumptions commonly include concepts such as starts and destinations of trips, commuting, and the dependency on a road network. Popular algorithmic generators in research are BerlinMOD, Brinkhoff's generator, and SUMO. As the latter two do not model entities over more than the duration of a single trip, the nature of the data they produce is not the same as that of real data. For the remainder of the paper, we therefore focus on BerlinMOD as the algorithmic generator.

As a second and novel approach to generating geolocation data, Artificial Neural Networks (ANN) are considered in this paper. ANNs have grown increasingly popular over the past years and they have been successfully applied to various kinds of problems. Two types of ANN are evaluated for generating geolocation data in this paper: Recurrent Neural Neural Networks because of the demonstrated ability to generated sequence data and Generative Adversarial Networks because they were specifically developed to generate data that is indistinguishable from real data. The particular architectures developed for this paper are introduced in sections 3.1 and 3.2, respectively.

3.1 Recurrent Neural Networks

The particular Recurrent Neural Networks (RNN) architecture used in this paper builds on an architecture for text generation presented by Martin Gorner of *Google* [Go17, Te17]. The architecture consists of a stack of Gated Recurrent Unit (GRU) cells, followed by time-distributed fully-connected layers (see Figure 1). Time-distributed here means that there are separate fully-connected layers following each of the GRU cells at each time step. Therefore, the output of the network is a sequence of 3-tuples. Different numbers of stacked GRU cells, fully-connected layers, and both their sizes may be chosen.

The given RNN architecture requires the training data to be fed into it in a way that respects that training is run in batches and on sequences of geolocation data, that a validation is run at the end of each epoch of training, and that the RNN's state remains persistent. For the specifics and a formal description of this formatting of the data please refer to the

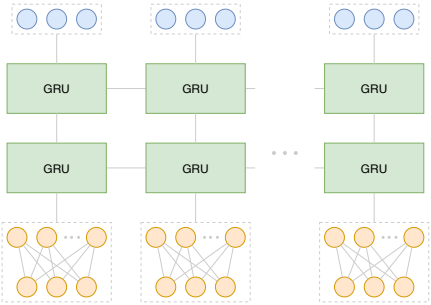


Fig. 1: RNN architecture

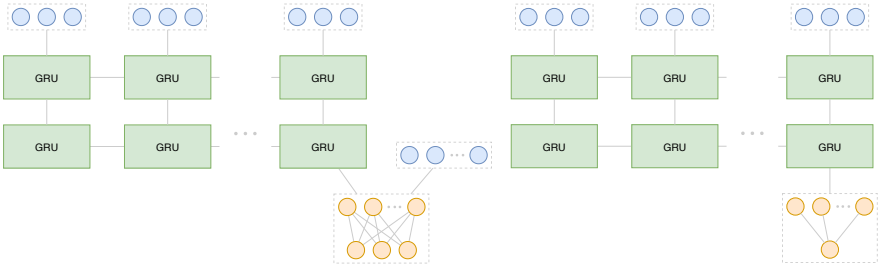
thesis [Ka19] this paper is based on. This thesis also goes into more details on the exact parameters, such as loss functions and optimisers, we used.

3.2 Generative Adversarial Networks

Generative Adversarial Networks (GAN) [Go14] use an *adversarial model* – a setup where two ANNs compete against each other in becoming very good at their respective tasks – in order to train an ANN that can generate data that, in the ideal scenario, is new but indistinguishable from real data. The general idea is a setup of two ANNs, a *generator* and a *discriminator*. The discriminator's task is to tell generated fake data from real data, and the generator's task is to generate fake data based on a random seed that the discriminator cannot tell apart from real data.

The particular GAN architecture used in this paper consists of both a recurrent generator and a recurrent discriminator. RNN were chosen for the GAN architecture as well because the task is to generate sequences. The generator uses a sequence of stacked GRU cells where only the output of the last stack is concatenated with a random input. The concatenated vectors are fed forward to a stack of fully-connected layers, the last of which must always have three neurons. The height of both the GRU cell stack and the fully-connected stack may be chosen freely. The generator's architecture is illustrated in Figure 2a. A similar architecture is used for the discriminator. Again, a stacked sequence of GRU cells is used and only the last cell's output is fed forward into a stack of fully-connected layers. The last of the fully-connected layers of the discriminator must have only a single neuron. See Figure 2b for an illustration of the discriminator network used in this paper.

For training the GAN, the real data needs to be provided in the correct format to ensure the coherence of the networks' states. For details on the particular formatting of the data, including formal descriptions, we, again, refer to the thesis [Ka19] this paper is based on. The latter thesis also gives information on the training process, loss functions, and optimisers we used.



(a) Generator (b) Discriminator

Fig. 2: RNN architectures used in the GAN setup

4 Interrogations

The purpose of generated geolocation data is to be used instead of real recorded geolocation data when no suitable real data is available. Ideally, generated data should therefore be indistinguishable from actual real data, i.e. it should be fully realistic. However, defining realism for geolocation data is not trivial. A multitude of complex factors influences real geolocation data and the general definition of realism on geolocation data is equally ill defined as the concept of intelligence. To the best of our knowledge, there exists no formal way of fully grasping either of these matters in its entirety.

The determination of intelligence has famously been approached using the Turing Test [Tu50]. The idea is that a machine is considered intelligent, if an interrogator cannot point out the machine after asking questions in a conversation with both a human and the machine. A common setup for the test is to measure the proportion of interrogators who were unable to tell which conversational partner is the machine. Thereby, this approach circumvents the need for a thorough definition of intelligence. Setups similar to the Turing Test have been proven to work well in other fields such as distinguishing paintings created by GAN architectures from those created by human artists [El17].

Mapping the Turing Test approach to geolocation data sets, the interrogator is given a real and a generated data set – not knowing which of the two is which – and the task to figure out which of the two data sets is generated. In place of a conversation, the interrogator may perform different kinds of data analysis, referred to as *questions*, on both data sets. While we as the interrogators do know which data sets are real or fake, we attempt to recreate the way an uninformed interrogator would argue.

4.1 Real Data Sets

For the purpose of a Turing Test-like interrogation we compared four real recorded data sets to data sets generated by BerlinMOD and our ANN generators. We used the same

data sets to train the ANN generators. All real data sets and their properties are given in Table 1. The chosen data sets are to our knowledge the largest freely available geolocation data sets currently used in research. The Mobile Data Challenge [Ki10, La12] (MDC) and GeoLife [ZXM09, ZXM08, ZXM10] data sets are both based on people. Because data sets based on people are rare, we also include the Cabspotting [PSDG09] and T-Drive [Yu11, Yu10] data sets, which are taxi-based.

	MDC	**GeoLife**	**Cabspotting**	**T-Drive**
Location	Lake Geneva region	Beijing & world	San Francisco	Beijing
Entities	185 users	182 users	536 taxis	10,357 taxis
Duration	1.5 years	5 years	30 days	6 days
Records	13,678,618	24,876,978	11,219,955	17,662,984

Tab. 1: Properties of the real data sets

4.2 Data Set Generation Setup

We generated the artificial data sets to have the same basic properties, e.g. city, number of users, and duration, as their real counterparts. For generating data using BerlinMOD we used data from OpenStreetMap [Op17] as the input for the road network.

	rnn	***gan_gen***	***gan_dis***
GRU	32	32	16
Fully-connected	3	3	1

Tab. 2: ANN configurations with each number in the GRU and Fully-connected row representing a layer and giving its size

All ANNs were trained on the respective real data sets. We tried different configurations of the presented architectures, but all produced similar results. We therefore only present the results of the configurations from Table 2 using a batch size of 64, a sequence length of 100, *mean absolute error* (RNN) and *binary cross entropy* (GAN) loss functions, and the *Adam* optimiser [KB14]. At the start of the generation we use *Kernel Density Estimation* (KDE) [Ro56, Pa62] to create distributions from the real data sets, from which we sample latitude and longitude of the initial records as well as relative start and end times of each user's data. The start position KDE is fitted on all positions in the full original data set using a bandwidth of 0.00003. The KDE for the relative start time is fitted only on a reduced data set of 50 users and 1 week (see Section 4.3). A bandwidth of 0.03 is used for fitting the KDEs to relative start times and durations. Sampling of relative start times and durations is repeated until the relative start time lies within the interval $[0, 1[$ and the sum of relative start time and duration is within the interval $]0, 1]$, in order to ensure that only data within the given week is generated. We then convert relative times to real times before feeding them to the ANNs to generate new records starting from the initial record for as long as the generated records' timestamps are within the sampled duration. Because the ANNs generate time deltas δt instead of absolute times, we constrain each time delta to be at least 1 second

for RNN and 90 seconds for GAN in order to make sure that the generation eventually terminates. A larger lower bound is needed on the GAN generation because the trained GANs tend to produce small δt.

4.3 Interrogation Setup

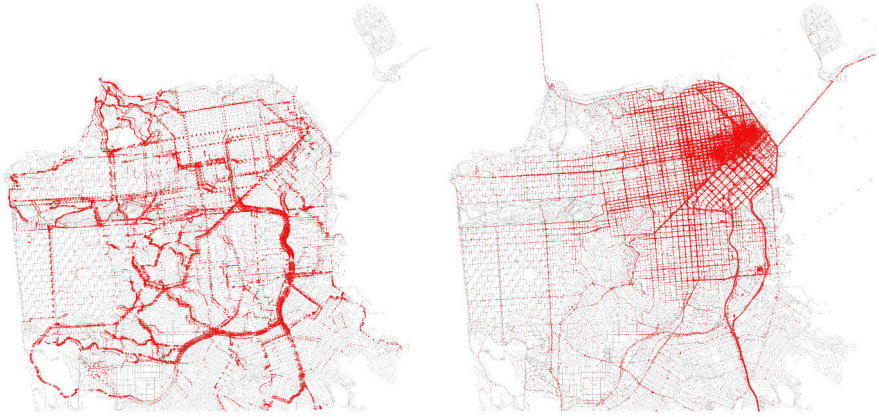
For this evaluation, all data sets – real and generated ones – are limited to 50 randomly chosen users and the time frame of the busiest week (Monday to Sunday for all but T-Drive) in order to keep computation times manageable. An exception is made on the T-Drive data set. Because this data set does not cover a full Monday to Sunday week, its entire time span is considered.

For the sake of brevity, we present only 4 of the 12 interrogations we conducted, two on data sets generated by BerlinMOD and one on each of the ANNs we presented (see [Ka19] for the complete observations of all interrogations). We assume that the intent is to generate a data set of the same nature – people- or taxi-based – as the real counterpart and that the interrogator is aware of this information. We also devised four questions our interrogator will ask. However, our interrogator is free to make a final decision after any number of questions.

The first question we devised is a simple plot of all records as red dots on a map of the respective city's streets, allowing the interrogator to quickly spot irregular behaviour. We call this a *map overview*. The second question our interrogator can ask is a histogram of the number of records over bins of all 168 hours of the relevant week. We hope to observe behaviour such as commuting in this histogram we refer to as *traffic*. Our third question we call *speeds*. This question constitutes a histogram of the speeds measured between two consecutive records for speeds from 0 to 150 km/h. We hope to be able to spot unlikely speeds in a data set this way. As the fourth and last question we devised, we plot the records of a single user coloured according to their time of recording on a map. This question may give clearer insights into user behaviour than the map overview question. We refer to this question as *single user*.

4.4 BerlinMOD versus Cabspotting

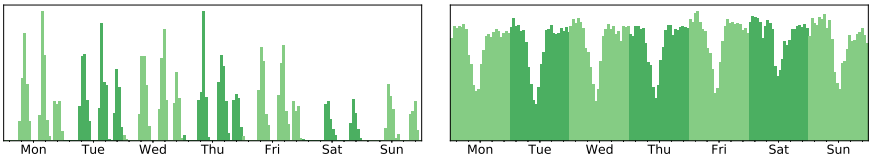
Starting with the data set generated by BerlinMOD based on the Cabspotting data set, we first consider a map overview (see figures 3a and 3b). Both data sets mostly follow the street network. Only a few of the records are not on the streets, which can be explained as either disturbance caused by the GPS positioning systems or users entering a building. The records which leave the streets display different behaviours in both data sets. In data set 2 those records are placed arbitrarily, with some even being in the sea. In data set 1, on the other hand, off-street records remain close to the streets. This may be a soft argument to consider



(a) BerlinMOD (data set 1) (b) Real (data set 2)

Fig. 3: Map overview of data sets on San Francisco, USA

data set 1 to be artificial as it gives the impression of being perfect data that was simply disturbed slightly after it was generated. Furthermore, in data set 2 there is a single area, where records are concentrated densely, whereas data set 1 is very focused on the street network only. Given that Cabspotting is a taxi-based data set, it can be argued that data set 1 is more realistic because taxis are not likely to leave the streets. However, the argument could also be made that most cities have a main area of interest, where traffic accumulates, and that such an area can only be observed in data set 2. Going by this question, no clear decision can be made regarding which of data sets is the artificial one.

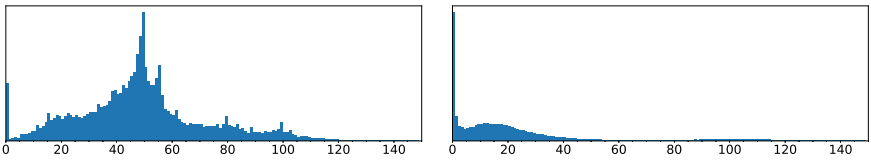


(a) BerlinMOD (data set 1) (b) Real (data set 2)

Fig. 4: Number of records during each hour in data sets on San Francisco, USA

Next, we ask the question of traffic (see figures 4a and 4b). data set 2 shows that there is less traffic during the midday on each day. However, one would intuitively expect there to be the least amount of traffic at night. Furthermore, all days of the week show the same behaviour and, except for Sunday, similar amounts of traffic. Intuitively, one would expect the days of the weekend to behave differently from work days and be less busy altogether. The first immediately visible property of data set 1 is that there are periods of no traffic. This may be possible in data sets of few users, but one would intuit that in 50 taxis, there is always at least one moving. Otherwise, traffic seems to follow a commuting pattern. Taxis travel in the morning hours and in the afternoon, and smaller spikes occur in the evenings. On

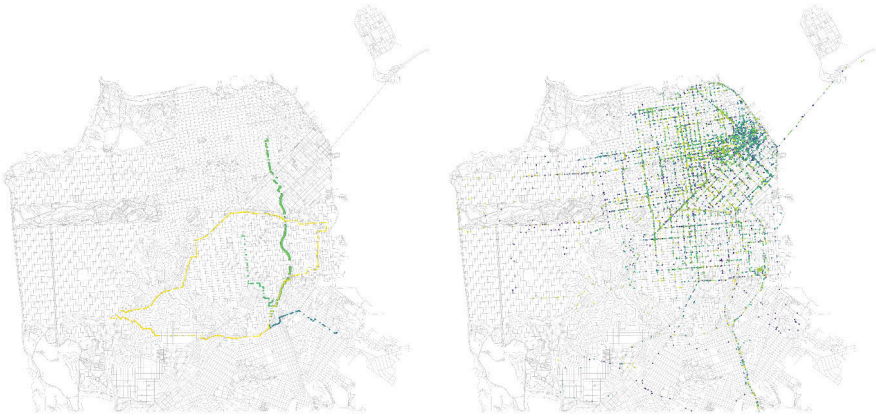
the weekend, taxis log fewer records than on weekdays and at different times of the day. Assuming that taxis are used by people for their commute and travels to free times activities, these patterns make sense. Based on these observations, it is hard to make out the artificial data set. Only the times of no traffic in data set 1 make that data set seem slightly strange and possibly artificial.



(a) BerlinMOD (data set 1) (b) Real (data set 2)

Fig. 5: Histograms of user's speeds in data sets on San Francisco, USA

So far a strong decision is not possible, so we move on to the question of speeds (see figures 5a and 5b). In data set 2 users move at less than 40 km/h most of the time, which seems about right for taxis in city traffic, but there is also a smaller cluster surrounding 100 km/h, presumably caused by taxis travelling on motorways that cut through the city. There is also a spike near 0 km/h which we argue could be caused by taxis waiting for passengers. In data set 1 speeds peak at around 50 km/h – a common speed in cities – and drop towards 0 km/h and 120 km/h, meaning there are taxis that travel slowly, for example in traffic jams or on small streets, and some that travel fast, for example on motorways. The smaller spike at 0 km/h can be explained as taxis waiting for passengers. Summarising the question of speeds, neither of the two data sets can be pointed out as generated.



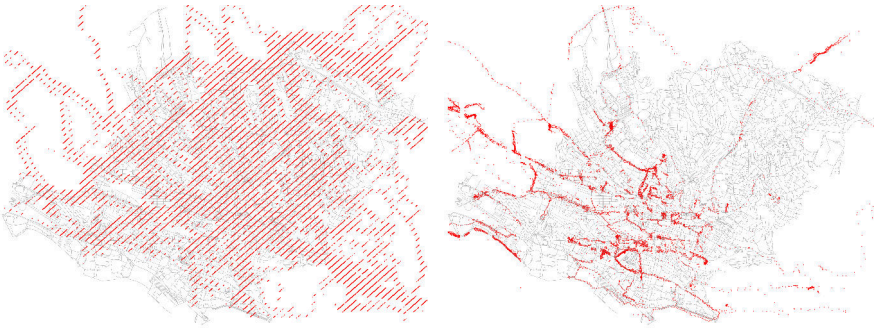
(a) BerlinMOD (data set 1) (b) Real (data set 2)

Fig. 6: Data of single users on San Francisco, USA

Because the previous question did not yield a clear indication, we decide to look at single trajectories (see figures 6a and 6b). An immediately obvious difference between both data sets is that the user in data set 2 went to many different places, whereas the user from

data set 1 looks orderly and only visited a handful of places mostly travelling along the same routes. In a people-based data set the latter behaviour would make sense with people frequently commuting between their homes and work places on the same routes, but given that Cabspotting is a taxi-based data set, this last question gives another soft indication towards data set 1. As taxis transport many different passengers to and from many different locations, they are likely to visit many different places over the course of one week.

4.5 BerlinMOD versus MDC



(a) BerlinMOD (data set 3) (b) Real (data set 4)

Fig. 7: Map overview of data sets on Lausanne, Switzerland

Comparing a data set generated by BerlinMOD to the MDC data set, we start with a map overview (see figures 7a and 7b). In both data sets, users follow the roads most of the time and both data sets appear to have a certain degree of disturbance in them, something that is natural to GPS positioning systems. However, the disturbance in data set 3 occurs along the same axis for all records, which we do not expect to observe in real data. We therefore conclude that data set 3 is generated and do not proceed to the next question.

4.6 RNN versus Cabspotting

Comparing the data set generated by our RNN architecture to the Cabspotting data set, we once again first ask the map overview question (see figures 8a and 4b). For the reasons given in Section 4.4, data set 2 does not give any indicators that it is artificial. The data in data set 5, however, appears to evolve only around a single point on the map, and when moving away or towards this point the users appear to do so with absolute disregard for the streets. This feature clearly does not resemble what one might expect a real data set to look like, hence we conclude already that data set 5 was artificially generated. We ask no further questions.

(a) RNN (data set 5) (b) GAN (data set 6)

Fig. 8: Map overview of data sets on San Francisco, USA

4.7 GAN versus Cabspotting

Moving on to the data set generated by our GAN architecture based on the Cabspotting data set, we start with the map overview question (see figures 8b and 4b). We immediately decide that data set 6 is artificial because the features of data set 2 can be argued to be realistic as already done in Section 4.4 and data set 6 displays a similar strange behaviour as data set 5 in Section 4.6. Taxis in data set 6 do not move along the streets but instead follow a circular path into a small area, where they remain mostly stationary. This area is shared among all taxis. We would not expect such behaviour to be present in real geolocation data and therefore decide that data set 6 is artificial without asking further questions.

4.8 Conclusion of Experiments

From the above experiments as well as the other experiments we conducted, we conclude that in only 1 out of 12 cases a generator was able to fool the interrogator. In this particular case of BerlinMOD generating a data set based on GeoLife, the sparse nature of the real GeoLife data set made the latter seem artificial in comparison to the dense data set generated by BerlinMOD. Other data sets generated by BerlinMOD also did well. Most times only details revealed their artificial nature after asking all four of our questions. Data sets generated by either of our ANN architectures did badly. In every one of the interrogations conducted on them, they could be pointed out as artificial immediately after asking the very first question of a map overview. On all four surveyed real data sets we observed the problems to be the same. Road networks appear to be ignored by the ANN generators in all cases, always producing similar circular movement which accumulates on one or a few positions on the map.

5 Conclusion

Concluding this study, we find that data generated by any of the three surveyed generators cannot replace real data. In all data we generated there remains enough evidence within recorded and generated data sets to tell them apart, which means that some of their properties regarding realism, which potentially have an influence on a prospective use case, do differ. Of the presented generators, BerlinMOD came the closest to producing a realistic data set. In most questions of the interrogation, data generated by BerlinMOD looked realistic at first glance, often only giving away its generated nature through detailed features of the data. Our generators based on ANNs did not do well. The data they generate is profoundly unrealistic and not at all suitable for testing algorithms. We therefore advise against using either one for testing and training algorithms that work on geolocation data. When the results are not required to be fully reliable, data generated by BerlinMOD may be used.

One might argue, as we have seen in the experiments ourselves, that a lack of quality in the real data can cause this method to fail as features of realistic data, such as sparsity or outliers, can fool interrogators depending on their expertise. However, because such effects are commonplace in real data, all discrimination methods do have to deal with the effects. Our Turing-based method of discriminating between real and generated data sets has otherwise proven to work well. We showed that our method functions well as a means of evaluating the realism of generated data while eliminating the need for a thorough definition of realism. The method we presented has potential for use on other kinds of data as it is domain-independent.

Opportunities for future work lie in devising further questions to ask on a data set. Such questions may be of use for evaluating more powerful generators. As an alternative to the presented approach based on the Turing Test, future work might also consider using ANNs as the means of evaluating the realism of generated geolocation data sets. Furthermore, our method based on the Turing Test may be applied to other types of data, of which in the age of *Big Data* there are certainly many. Training more capable ANN generators for geolocation data is also an intriguing idea for future work. ANNs could potentially also be used in the field of Trajectory Anonymisation by training them on a single real data set and then using them to produce a new data set with the same general features, but with users that behave slightly differently from the real users, thereby protecting the privacy of the real users.

Acknowledgements

(Portions of) the research in this paper used the MDC Database made available by Idiap Research Institute, Switzerland and owned by Nokia. Map data copyrighted OpenStreetMap contributors and available from `https://www.openstreetmap.org`.

References

[Br03] Brinkhoff, T.: Generating traffic data. IEEE Data Engineering Bulletin, 26:19–25, 2003.

[DBG09] Düntgen, C.; Behr, T.; Güting, R. H.: BerlinMOD: A benchmark for moving object databases. The VLDB Journal, 18:1335–1368, 2009.

[El17] Elgammal, A.; Liu, B.; Elhoseiny, M.; Mazzone, M.: CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms. arXiv preprint, arXiv:1706.07068, 2017.

[Go14] Goodfellow, I. et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems. pp. 2672–2680, 2014.

[Go17] Gorner, M.: TensorFlow and Deep Learning without a PhD. https://www.youtube.com/watch?v=fTUwdXUFfI8 (Accessed: 12th April 2019), 2017.

[Ka19] Kaiser, J.: A study of generated versus recorded geolocation data. Research project thesis, Hamburg University of Technology, 2019. Available at https://www.sts.tuhh.de/pw-and-m-theses/2019/kaiser19.pdf.

[KB14] Kingma, D. P.; Ba, J.: Adam: A Method for stochastic optimization. arXiv preprint, arXiv:1412.6980, 2014.

[Ki10] Kiukkonen, N.; Blom, J.; Dousse, O.; Gatica-Perez, D.; Laurila, J.: Towards rich mobile phone datasets: Lausanne data collection campaign. In: Proceedings of the ACM Int. Conf. on Pervasive Services. 2010.

[Kr12] Krajzewicz, D.; Erdmann, J.; Behrisch, M.; Bieker, L.: Recent development and applications of SUMO - Simulation of Urban MObility. Int. Journal On Advances in Systems and Measurements, 5:128–138, 2012.

[La12] Laurila, J. et al.: The mobile data challenge: Big data for mobile computing research. In: Proceedings of the Workshop on the Nokia Mobile Data Challenge, in Conjunction with the 10th Int. Conf. on Pervasive Computing. pp. 1–8, 2012.

[MBT11] Martinez-Bea, S.; Torra, V.: Trajectory anonymization from a time series perspective. In: IEEE Int. Conf. on Fuzzy Systems. pp. 401–408, 2011.

[Mo16] Mogren, O.: C-RNN-GAN: Continuous recurrent neural networks with adversarial training. arXiv preprint, arXiv:1611.09904, 2016.

[Op17] OpenStreetMap contributors: Planet dump retrieved from https://planet.osm.org. https://www.openstreetmap.org, 2017.

[Pa62] Parzen, E.: On estimation of a probability density function and mode. The Annals of Mathematical Statistics, 33:1065–1076, 1962.

[PSDG09] Piorkowski, M.; Sarafijanovoc-Djukic, N.; Grossglauser, M.: A parsimonious model of mobile partitioned networks with clustering. In: The First Int. Conf. on COMmunication Systems and NETworkS. pp. 1–10, 2009.

[Rh11] Rhee, I.; Shin, M.; Hong, S.; Lee, K.; Kim, S. J.; Chong, S.: On the Levy-walk nature of human mobility. IEEE/ACM Transactions on Networking, 19:630–643, 2011.

[RMC15] Radford, A.; Metz, L.; Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint, arXiv:1511.06434, 2015.

[Ro56] Rosenblatt, M.: Remarks on some nonparametric estimates of a density function. The Annals of Mathematical Statistics, 27:832–837, 1956.

[Te17] TensorFlow: Text generation using a RNN with eager execution. `https://www.tensorflow.org/tutorials/sequences/text_generation` (Accessed: 12th April 2019), 2017.

[Tu50] Turing, A. M.: Computing Machinery and Intelligence. Mind, 59:433–460, 1950.

[Yu10] Yuan, J. et al.: T-Drive: Driving directions based on taxi trajectories. In: Proceedings of 18th ACM SIGSPATIAL Conf. on Advances in Geographical Information Systems. pp. 99–108, 2010.

[Yu11] Yuan, J.; Zheng, Y.; Xie, X.; Sun, G.: Driving with knowledge from the physical world. In: Proceedings of the 17th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. pp. 316–324, 2011.

[Zh16] Zhao, K.; Tarkoma, S.; Liu, S.; Vo, H.: Urban human mobility data mining: An overview. In: 2016 IEEE Int. Conf. on Big Data. pp. 1911–1920, 2016.

[Zh17] Zhu, J. Y.; Park, T.; Isola, P.; Efros, A. A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE Int. Conf. on Computer Vision. pp. 2223–2232, 2017.

[ZXM08] Zheng, Y.; Xie, X.; Ma, W. Y.: Understanding mobility based on GPS data. In: Proceedings of the 10th Int. Conf. on Ubiquitous Computing. pp. 312–321, 2008.

[ZXM09] Zheng, Y.; Xie, X.; Ma, W. Y.: Mining interesting locations and travel sequences from GPS trajectories. In: Proceedings of the 18th Int. Conf. on World Wide Web. pp. 791–800, 2009.

[ZXM10] Zheng, Y.; Xie, X.; Ma, W. Y.: GeoLife: A collaborative social networking service among user, location and trajectory. IEEE Data Engineering Bulletin, 33:32–39, 2010.

Inductive Learning of Concept Representations from Library-Scale Corpora with Graph Convolution

Lukas Galke[1] Tetyana Melnychuk[2] Eva Seidlmayer[3] Steffen Trog[4] Konrad U. Förstner[5] Carsten Schultz[6] Klaus Tochtermann[7]

Abstract: Automated research analyses are becoming more and more important as the volume of research items grows at an increasing pace. We pursue a new direction for the analysis of research dynamics with graph neural networks. So far, graph neural networks have only been applied to small-scale datasets and primarily supervised tasks such as node classification. We propose to use an unsupervised training objective for concept representation learning that is tailored towards bibliographic data with millions of research papers and thousands of concepts from a controlled vocabulary. We have evaluated the learned representations in clustering and classification downstream tasks. Furthermore, we have conducted nearest concept queries in the representation space. Our results show that the representations learned by graph convolution with our training objective are comparable to the ones learned by the DeepWalk algorithm. Our findings suggest that concept embeddings can be solely derived from the text of associated documents without using a lookup-table embedding. Thus, graph neural networks can operate on arbitrary document collections without re-training. This property makes graph neural networks useful for the analysis of research dynamics, which is often conducted on time-based snapshots of bibliographic data.

Keywords: machine learning; representation learning; neural networks; graph mining

1 Introduction

The investigation of bibliographic data enables rich insights into research dynamics including knowledge generation and diffusion, convergence of distinct scientific areas and substitution of some scientific fields with converged domains. New valuable knowledge is produced within scientific communities through collaboration of multiple actors [PHW12, WJU07]. A collaboration between researchers from different scientific fields fosters the diffusion of knowledge of one domain into other fields. The intensification of such collaborations leads to blurring the boundaries between separate scientific fields and to emerging scientific disciplines [CBL10].

[1] ZBW – Leibniz Information Centre for Economics, Kiel and Hamburg, Germany l.galke@zbw.eu
[2] Kiel University, Germany melnychuk@bwl.uni-kiel.de
[3] ZB MED – Information Centre for Life Sciences, Cologne, Germany seidlmayer@zbmed.de
[4] ZBW – Leibniz Information Centre for Economics, Kiel and Hamburg, Germany shtrog@gmail.com
[5] ZB MED – Information Centre for Life Sciences, Germany foerstner@zbmed.de
[6] Kiel University, Germany schultz@bwl.uni-kiel.de
[7] ZBW – Leibniz Information Centre for Economics, Kiel and Hamburg, Germany k.tochtermann@zbw.eu

doi:10.18420/inf2019_26

Library-scale corpora of scientific publications hold a large potential for automated analyses of research dynamics. Machine learning techniques that benefit from large amounts of data are essential for studying research dynamics. A major challenge in analysis of research dynamics is to derive a meaningful similarity measure. So far, most existing approaches rely on text-based similarity, co-citation analysis [NMF17, URU10], or scientometric methods [Je16, JLC18]. In contrast, we exploit concept annotations, as present in corpora of (scientific) digital libraries to derive a similarity measure between concepts. We make use of machine learning techniques to learn a low-dimensional continuous vector, i. e., a *representation* for each concept, from which a similarity measure can be derived.

Problem Statement In a paper-concept graph, we study the novel problem of learning representations for featureless concept nodes from paper nodes that have textual features. We evaluate whether the resulting concept representations are *meaningful*, i. e. correspond to human judgements, and *useful* in terms of their performance in downstream tasks.

Formally, we operate on a graph $\mathcal{G} = (\mathbb{P} \cup \mathbb{C}, \boldsymbol{X}, \boldsymbol{A})$, whose N vertices are either paper nodes $\mathbb{P}$ or concept nodes $\mathbb{C}$. Textual features of paper nodes are encoded in $\boldsymbol{X} \in \mathbb{R}^{|\mathbb{P}| \times L}$, where L is the textual feature dimension. Concept nodes have no features. Edges are encoded in the adjacency matrix $\boldsymbol{A}$ such that $A_{ij} > 0$ when either two papers $i, j < N$ have at least one common author or a paper i is annotated with concept j. The task is to learn a parametrized function f_θ that maps paper $\boldsymbol{X}, \boldsymbol{A}$ to concept representations $C \in \mathbb{R}^{|\mathbb{C}| \times d}$ of size d. To enable a fair comparison between methods, we keep d fixed because larger representation sizes tend to lead to increased performance in downstream tasks [ERG19].

We call a method *transductive* if it relies on a static concept embedding given by a look-up table. A method is *inductive*, when the concept representation C can be derived solely from the input corpus $\boldsymbol{X}, \boldsymbol{A}$ without conducting further training.

In this paper, we propose to use graph convolution [KW16a, Hu18, CZS18] to tackle this problem. To enable unsupervised representation learning, we introduce a dedicated training objective. We compare the resulting concept representations to the ones of transductive DeepWalk and text-based latent semantic analysis [De90].

Our results show that the representations learned by graph convolutional networks are similarly useful and meaningful as the representations learned by DeepWalk [PAS14]. At the same time, our graph convolution approach has the advantage that it does not rely on a static concept embedding but rather learns a mapping from associated papers to concept representations. This turns graph convolution into a valuable approach for the analyses of research dynamics. A once-learned model can induce concept representations for any (sub-)set of annotated research papers such as annual snapshots. This is important for the analyses of research dynamics, which we consider as future work.

In summary, our contributions are: (1) We apply state-of-the-art graph neural networks on a dataset of 2.1M papers from the economics and business studies domain. (2) We introduce

a dedicated, reconstruction-based training objective that allows unsupervised representation learning of concept representation. (3) We show that the learned concept representations are similarly useful and meaningful as the ones of DeepWalk, while not depending on a static node embedding.

In the following Section 2, we give an overview of the related work. Then, we describe the employed methods in Section 3, before we outline the experimental setup in Section 4. We provide the results in Section 5, discuss them in Section 6, before we conclude.

2 Related Work

Salatino, Osborne, and Motta [SOM17] have shown that there is a strong correlation between the pace of collaboration and the emergence of new topics. The same authors have then developed an advanced clique percolation method [SOM18] to detect emerging topics at the early stages and evaluate against a synthetic ground truth. Wu et al. [WVC16] have studied the top 1% authors within the computer science domain and show that research topics are increasingly inter-related. Duvvuru, Kamarthi and Sultornsanee [DKS12] link keywords when they appear in the same scholarly article. The authors construct visual keyword maps that may aid identifying emerging research areas. He et al. [He09] propose to use citation data in conjunction with latent Dirichlet allocation to analyze topic evolution. Tseng et al. [Ts09] compare several methods to detect hot topics.

Several approaches have been proposed that are targeted specifically towards multi-relational graphs such as knowledge graphs or linked data [Bo13, So13, Ya14]. For homogeneous graphs, as faced in our context, the successful Word2vec algorithm [Mi13] has been transferred to graphs by sampling random walks, namely DeepWalk [PAS14]. Node2vec [GL16] generalizes DeepWalk and further analyzes how the window size affects capturing more structural or more semantic relationships. Yang, Cohen, and Salakhutdinov [YCS16] outline the difference between inductive and transductive learning settings and develop an approach that is suited for both cases. All previously described methods rely on look-up table embeddings and are, thus, not suited for inductive learning.

Numerous methods have recently emerged that generalize convolution to graphs [DBV16, KW16a]. In GraphSAGE [HYL17], the authors explore different aggregation functions and conduct experiments on representation learning in large-scale graphs by sampling adjacent nodes. Velickovic et al. [Ve18] suggest to incorporate an attention mechanism for neighbor aggregation. We refer to [Wu19] for a recent overview on graph neural networks.

3 Inductive Representation Learning with Graph Convolution

Graph convolution is an approach for graph-structured data that is capable of jointly exploiting textual and structural features. Approaches based on graph convolution yield

promising results on link prediction [KW16b], semi-supervised classification [KW16a], and representation learning [HYL17]. A benefit of graph convolution is the possibility to conduct inductive learning [YCS16, HYL17]. Inductive learning means that the textual features from paper nodes are aggregated to compose a representation of the featureless concept nodes. This property distinguishes this approach from other approaches that learn a static node embedding such as DeepWalk [PAS14] as well as TransE [Bo13] and their extensions. The inductive property allows computing concept representations on the basis of any subset of the data and also for entirely unseen data [GVS19].

To make use of graph convolution, we first embed the textual features into a lower-dimensional space by averaging word vectors $\boldsymbol{h}^{(0)} = \frac{1}{|x|} \sum_{t \in \boldsymbol{x}} \boldsymbol{W}^{(0)}_{t,:}$, where x are the words of a document. Subsequently, we make use of graph convolution to aggregate neighbor representations after a nonlinear transform. The representation of node i in layer l is defined as:

$$\boldsymbol{h}_i^{(l+1)} = \sigma\left(\sum_{j \in \mathcal{N}(i)} \frac{1}{c_{ij}} \boldsymbol{W}^{(l)} \boldsymbol{h}_j^{(l)} + \boldsymbol{b}^{(l)}\right)$$

where $\mathcal{N}(\cdot)$ refers to the set of adjacent nodes and σ is a nonlinear activation function. We follow [HYL17, Hu18, CZS18] and use mean aggregation $c_{ij} = |\mathcal{N}(i)|$. The weights $\boldsymbol{W}^{(0)}, \boldsymbol{W}^{(1)}, \boldsymbol{b}^{(1)}, \ldots, \boldsymbol{W}^{(k)}, \boldsymbol{b}^{(k)}$, with k being the depth of the network, are then optimized with respect to the training objective, which we describe in the following section.

Training Objective Unsupervised deep learning techniques exploit auxiliary objectives such as auto-encoding [BCV13]. An auto-encoding objective refers to the task of reconstructing the input. It may happen that the input-output space is high-dimensional. For instance, consider the vocabulary of all words. In these cases, normalizing across all output probabilities via softmax can become computationally expensive. Negative sampling [Mi13] approximates the softmax by sampling few negative outputs. The task is then to distinguish the true output among the negative samples. Due to its higher efficiency, negative sampling often yields higher effective scores than the exact computation of the softmax [Mi13].

In the graph domain, link prediction is a common choice for learning node representations. The representation is trained for predicting whether a link between two nodes exists. This can be regarded as auto-encoding the adjacency matrix [KW16a]. Also here, negative sampling can be employed to approximate the full softmax [PAS14, HYL17].

In our case, the goal is to learn concept representations for a controlled vocabulary. While our models deal with millions of research papers, the dimension of the controlled vocabulary is rather small with 5,688 concepts. We chose to use only concepts from the controlled vocabulary as optimization objective. Thus, we can afford to compute the full softmax over the concepts. We employ a linear decoder $g : \mathbb{R}^d \rightarrow \mathbb{R}^{|C|}$ that uses the final representation

of the graph convolutional network to reconstruct the respective concept. The loss function is the softmax over the concepts:

$$\mathcal{L}_{\text{rec}}(\boldsymbol{X}, \boldsymbol{A}, y) = -\log \frac{\exp g(f(\boldsymbol{X}, \boldsymbol{A}))[y]}{\sum_j \exp g(f(\boldsymbol{X}, \boldsymbol{A}))[j]}$$

where f is a graph convolutional encoder, j iterates through all concepts and $[\cdot]$ denotes index access. We sample a set of documents $\boldsymbol{X}$ which are connected over at most two hops to the true concept y. The graph convolutional encoder f then constructs a low-dimensional concept representation $f(\boldsymbol{X}, \boldsymbol{A})$, which is then used by g to reconstruct the true concept. Since g is discarded after training, we deactivate its bias term such that all information for prediction of the concept is drawn from the representation.

Neighbor Sampling and Skip Connections The originally proposed graph autoencoders [KW16b] and graph convolutional networks [KW16a] operate on the whole graph in each optimization step. Storing the dense adjacency matrix is, however, not an option when the dataset is of large scale. The authors suggest to construct mini-batches with adjacent nodes. However, the receptive field still grows exponentially with the number of layers. Hamilton et al. [HYL17] instead propose to subsample adjacent nodes such that the growth factor is constant. Unfortunately, subsampling does not guarantee convergence to the full graph convolution [CZS18]. A control variate approach has been proposed that provably converges to the optimal, full graph convolution solution with only two sampled neighbors [CZS18]. The authors propose to keep track of past activations to incorporate the difference into the forward propagation path. Huang et al. [Hu18] propose a sampling approach that makes use of skip-connections to preserve second-order connections throughout the sampling process. We adopt these advances and employ a graph convolutional network with control variate sampling and skip-connections. We do not insert self-loops, such that the inductive property is retained. After training, we use all neighbors for creating the final representations.

4 Experimental Setup

In the following Section 4.1, we will describe the characteristics of the dataset and the processing of textual and structural features. We described the employed baselines in Section 4.2 and denote the selected hyperparameters in Section 4.3, before we describe the evaluation measures in Section 4.4.

4.1 Dataset

The EconBiz dataset comprises more than 11M records describing scientific publications from the economics and business studies domain. About 5.8M of these records are well

described by a controlled vocabulary and are used for our investigations. We filter these publications for English language and for annotations from the polyhierarchically-organized Standardthesaurus Wirtschaft[8]. These annotations are created by professional subject indexers. The resulting subset consists of 2.1M publications along with 5,688 subjects from the controlled vocabulary. As we focus on concept representations, we collate the authorship edges between authors and papers. We create an edge between two papers if the two papers have an author in common. This effectively enlarges the size of the receptive field of the models by one hop. This holds not only for graph convolution, but also for DeepWalk.

We consider the titles of the documents as textual features. We have shown in prior work that using titles is competitive [Ge17] to full-text data for multi-label classification. When the amount of available title data exceeds the amount of full-text data, classifiers based on title data can even outperform classifiers based on full-text data [MGS18]. Thus, we employ the larger amount of available title data. For preprocessing, we remove punctuation and other non-alphanumeric characters, lowercase the text, and remove English stop-words. We compose a vocabulary of the 50,000 most-common words.

4.2 Baselines: LSA and DeepWalk

As baselines, we consider DeepWalk [PAS14] as a representative for a purely structural approach to graph representation learning along with latent semantic analysis [De90] as a well-known text-based approach for document-level similarity.

Latent semantic analysis [De90, MRS08] (LSA) is a technique to embed text documents into a lower dimensional space. The key idea of LSA is to factorize the term frequency–inverse document frequency [SB88] weighted term-document matrix. We apply LSA on the titles of the research papers [Ge17]. We employ truncated singular value decomposition to embed each document in a low dimensional vector space. Finally, we compute the centroid for each concept across those documents that are annotated with the respective concept.

DeepWalk [PAS14] is an approach for learning node embeddings in graph-structured data. The algorithm samples random walks through the graph structure. For each node in the path, its embedding is used to predict its predecessors and successors along the random walk. The embedding is initialized randomly and updated according to hierarchical softmax loss.

4.3 Hyperparameters

LSA uses 5 epochs for singular value decomposition of the term-document matrix. For DeepWalk, we generate 40,000 random walks for each concept node with a walk length of 3. We then run skip-gram optimization with window size 3 for 5 epochs over the generated

[8] http://zbw.eu/stw

random walks. The graph convolutional network uses two graph convolution layers. The text embedding size is 256 along with 128 hidden units and 128 output units corresponding to the representation size. We create mini-batches over concept nodes and sample 10 neighbors for each of the two hops. We run one sampling step per concept over 400 epochs. We use ReLU activation function and dropout [Ni14] with probability 0.5 within the GCN layers. We optimize the training objective via Adam [KB14] and an initial learning rate of 0.001. For a fair comparison, we fix the representation size to 128 for all models. Furthermore, the parameters are set such that both GCNs and DeepWalk are given the same number of sampled documents. We select a window size of 3 for DeepWalk such that the number of considered hops is the same as for GCNs.

4.4 Evaluation measures

To evaluate the resulting representations, we compare the performance on two downstream tasks: classification and clustering. For this purpose, we construct a dataset that maps each concept to its respective subthesaurus. The models have never seen the underlying concept hierarchy. As the thesaurus is organized in a polyhierarchic way, we use only those concepts, which belong to exactly one subthesaurus. We are left with 3,113 concepts and 7 classes.

Supervised Clustering We conduct a clustering on top of the learned concept representations with k-Means and k-Means++ as initialization strategy. We fix the number of clusters to 7 corresponding to the number of classes. We evaluate the supervised clustering metrics homogeneity, completeness, and V measure [RH07], as well as the adjusted rand index [HA85]. Homogeneity yields values between 0 and 1, which assess to which extent the clusters cover data points of the same class. Completeness is equivalent to homogeneity but switches the true and the predicted labels. V measure is the harmonic mean between homogeneity and completeness. The adjusted rand index is bounded between -1 and 1 and symmetrically assesses the similarity of a clustering result with the class labels. It is permutation-invariant and adjusted against chance. We report the mean scores of 100 k-Means runs for the raw concept vectors and L2-normalized concept vectors.

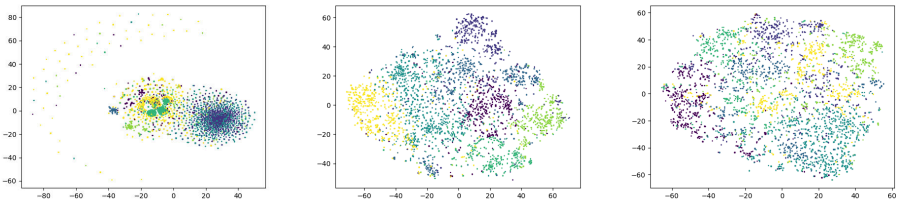
Unsupervised Clustering To gain more insights on the clustering tendency of the learned representations, we conduct a further unsupervised clustering experiment. Now we set the number of clusters to 101 corresponding to the number of top-level concepts across the 7 subthesauri. We evaluate the unsupervised clustering metrics silhouette coefficient [Ro87] and the Calinski-Harabasz criterion [CH74]. The silhouette coefficient is bounded between -1 and 1 and gives the ratio between intra-cluster distances and the pairwise distances to data points of the nearest cluster. The Calinski-Harabasz criterion compares the intra-cluster variance against the global between-cluster variance in distances. We also report these unsupervised clustering metrics for the supervised clustering experiments described above.

Classification We evaluate the performance in a downstream classification task. We use the L2-normalized learned concept representations as input and the corresponding subthesaurus as class label. As a common classifier we employ a support vector machine with linear kernel. We conduct a ten-fold cross-validation and report the mean accuracy.

5 Results

Tab. 1: Silhouette score (S), Calinski-Harabasz score (CH), homogeneity (H), completeness (C), V measure (V), and adjusted rand index (ARI) of clustering results on the learned concept representations for LSA, DeepWalk, and GCNs. We provide the mean of 100 k-Means runs with 7 clusters on 3,113 concept representations. Higher is better.

Model	Norm	S	CH	H	C	V	ARI
Random	None	0.0062	13.83	0.0032	0.0030	0.0031	0.0000
Random	Unit L2	0.0062	13.92	0.0033	0.0031	0.0032	0.0001
LSA	None	-0.0207	53.45	0.0030	0.0071	0.0042	-0.0041
LSA	Unit L2	**0.1284**	96.44	0.0022	0.0025	0.0023	-0.0009
DeepWalk	None	0.0194	124.80	0.2165	0.2496	0.2318	0.1852
DeepWalk	Unit L2	0.0670	131.18	**0.2930**	**0.2810**	**0.2869**	**0.1981**
GCN	None	0.0667	171.13	0.1845	0.1761	0.1802	0.1178
GCN	Unit L2	0.0823	**193.64**	0.1992	0.1891	0.1940	0.1423



(a) (b) (c)

Fig. 1: t-SNE visualization (perplexity=30) of the L2-normalized learned representations by LSA (a), Deepwalk (b) and GCN (c). The colors correspond to one clustering result with 7 clusters.

Table 1 shows the results for the supervised clustering task. For the supervised clustering metrics, the Deepwalk representations achieve the highest scores of 0.2930 homogeneity, 0.2810 completeness, 0.2869 V measure and adjusted rand index 0.1981. The scores of GCN representations are behind with a margin of 0.1 V-measure and 0.08 adjusted rand index. LSA representations yield the highest silhouette coefficient while GCN representations yield the highest Calinksi-Harabasz score. We provide a visualization of one clustering run in Figure 1.

Tab. 2: Mean silhouette score and Calinski-Harabasz score across 100 k-Means runs for the unsupervised clustering experiments with 101 clusters on learend representations of 5,688 concepts.

Model	Norm	Silhouette	Calinski-Harabasz
LSA	None	0.0543 (SD: 0.01)	32.14 (SD: 0.26)
LSA	Unit L2	0.0909 (SD: 0.01)	25.05 (SD: 0.13)
DeepWalk	None	0.0383 (SD: 0.00)	52.50 (SD: 0.34)
DeepWalk	Unit L2	0.0688 (SD: 0.00)	53.31 (SD: 0.13)
GCN	None	0.0721 (SD: 0.00)	72.78 (SD: 0.24)
GCN	Unit L2	**0.1005 (SD: 0.00)**	**84.88 (SD: 0.20)**

The results for the unsupervised clustering tasks with 101 clusters are shown in Table 2. Here, the GCN representations lead to the highest silhouette and Calinski-Harabasz scores of 0.1005 and 84.88, respectively.

In Table 3, we show the nearest-concepts for manually-selected concepts. The LSA representations fail to yield consistently explainable responses. For example, *Tax* is closest to *Rehabilitation hospital* and *Abortion*. The responses by GCN's and DeepWalk's representations are similarly acceptable: the closest concepts to *Tax* are in both cases all related to taxes. In case of *Germany*, DeepWalk returns other European countries but also *Comparison*. GCN yields parts of Germany along with Western Europe and Austria. We note that also linear relationships are resembled by both GCN and DeepWalk. For instance, the sum of the *Tax* vector and the *Theory* vector has *Theory of Taxation* among the two nearest concepts in the representation space. Similarly, the addition of *Economic growth* and *Theory* leads to having *Growth Theory* among the top two nearest concepts.

Table 4 shows the results for the downstream classification task. The non-normalized GCN representation achieves the highest classification accuracy of 68%. The highest scores for LSA and DeepWalk are 23% and 67%, respectively.

6 Discussion

Our results show that the representations of graph convolution are comparable to the ones of DeepWalk. While DeepWalk has lead to higher scores in the clustering task, GCN's representations have lead to higher scores in the classification downstream task. By inspecting the representations with nearest neighbor queries, we could observe that both DeepWalk and GCN correspond to human intuition, while LSA falls behind.

We have further analyzed the usefulness of the learned representations in an unsupervised clustering task with 101 clusters, enforcing a more fine-grained setting. In this setting, the

Tab. 3: Most similar concepts according to learned representations of LSA, DeepWalk, and GCN. The responses are ordered by descending cosine similarity to the vector of the query concept. A plus in the query column indicates that we use the sum of two concept vectors as query.

Query	LSA	DeepWalk	GCN
Economic growth	Management information system	Economic adjustment	Stages of growth model
	Tobacco	Economic policy	Growth policy
	Internet Usage	Growth policy	Resource wealth
	Eurobond	Economic development	Kuznets curve
	Automobile engine	Economic reform	Export-led growth
Tax	Rehabilitation hospital	Fiscal administration	Tax policy
	Abortion	Tax system	Tax system
	Biodiversity	Tax policy	Tax reform
	Financial statement analysis	Sales tax	Taxation procedure
	Association agreement	Tax reform	Tax burden
Germany	Debt crisis	Italy	East Germany
	Mesoeconomics	France	Austria
	Population policy	Comparison	West Germany
	Complaint management	Netherlands	Lower Saxony
	Unemployment theory	Austria	Western Europe
Vehicle	Pigouvian tax	Transport research	Sustainable mobility
	Cargo shipping	Transport economics	Passenger transport
	Cyclical unemployment	Waste treatment	Freight transport
	Wage subsidy	Battery	Major electrical appliances
	Financial Statement analysis	Microsystems	Traffic
Tax + Theory	Tax	Tax	Theory of taxation
	Theory	Theory of taxation	Theory
	Financial statement analysis	Tax system	Second best
	Nursing profession	Capital income	Optimal taxation
	Rehabilitation hospital	Public economics	Welfare economics
Economic growth + Theory	Economic growth	Economic growth	Growth theory
	Banking services	Growth theory	Neoclassical growth model
	Producer cooperative	Economic model	Unbalanced growth
	Licence	Theory	Balanced growth
	Laboratory	Endogenous growth model	Functional income distribution

GCN's representations have yielded the highest silhouette coefficient and Calinski-Harabasz score.

The strong performance of the DeepWalk is to some extent surprising, as it does not use any textual features but only relies exclusively on the structure of the author-paper-concept graph. This, however, confirms the claim of the original work [PAS14] that meaningful node embeddings can be derived without using node attributes.

There is no ground truth for pairwise similarity between concepts. We could therefore evaluate only a small subset of nearest-concept queries manually. We, however, did create a dataset which maps each concept to the respective subthesaurus. The hierarchical

Tab. 4: Downstream classification performance with 3,113 concepts and 7 classes. We list mean and standard deviation from a 10-fold cross-validation using a linear SVM classifier.

Model	Norm	Accuracy
LSA	None	0.2345 (0.00)
LSA	Unit L2	0.2181 (0.02)
DeepWalk	None	0.6625 (0.04)
DeepWalk	Unit L2	0.6708 (0.03)
GCN	None	**0.6813 (0.03)**
GCN	Unit L2	0.6496 (0.03)

relationships were never presented to the models, but only used for evaluation. The assumption is that the learned representation should allow distinguishing the concepts on a very broad level such as "Economics", "Business economics", "Geographic Names". A limitation of our study is that this categorization could be too broad to fully assess similarity among concepts. Constructing a more fine-grained evaluation set is challenging because the underlying thesaurus is polyhierarchic, i.e., a concept can have multiple broader concepts. Our subthesauri-based evaluation set uses only concepts that belong to exactly one subthesaurus, despite following all paths upwards in the hierarchy, which renders it well-defined, even in the polyhierarchic case.

We have applied our concept representation learning method to a large-scale dataset with 2.1M publications from the economics and business studies domain. Our approach can be transferred to any other dataset that is annotated with concepts. These concepts may come from a controlled vocabulary as in our case but free-text author keywords can be used instead. When scaling the number of concepts up, it can become necessary to switch from softmax training to a negative sampling approximation. Our model is flexible in the sense that it allows incorporating further edges such as the broader and narrower connections between the concepts. For now, we held out these connections for evaluation purposes.

The inductive property of the graph convolution approach enables us to map any set of annotated papers to concept representations without retraining. We can incrementally update the concept representations in a time-dynamic setting. This is important because fine-tuning pretrained, non-inductive, embeddings can be challenging: there are many options for weighting between the old embedding and the updates. We envision that this property will be crucial for analyses of the dynamics within and across research fields in our future work.

7 Conclusion

We conclude that the representations learned by graph neural networks are comparable to the ones learned by DeepWalk. Graph neural networks can induce representations for the

featureless concepts from the titles of associated research papers. To make graph neural networks applicable to library-scale bibliographic corpora, we have introduced a specific training objective for learning concept representations.

We have thoroughly analyzed the learned representations by conducting supervised and unsupervised downstream tasks. Furthermore, we have manually inspected the representations by conducting nearest neighbor queries. We have found that the nearest concepts are useful in downstream tasks and meaningful for humans, even in cases, where the vectors of two concepts are summed up.

Our findings suggest that concept embeddings can be solely derived from the text of associated documents without using a lookup-table embedding. In future work, we plan to make use of further structural features such as concept hierarchies. We further plan to use graph neural networks for dynamic research analyses based on annual snapshots of research papers. By analysing the trajectories, we can then make claims about the convergence and divergence of research areas.

Source Code: `github.com/lgalke/INFORMATIK2019-concept-representation-learning`

Acknowledgment: This work was supported by BMBF within the programme "Quantitative Wissenschaftsforschung" under grant numbers 01PU17013A, 01PU17013B, 01PU17013C.

Bibliography

[BCV13] Bengio, Yoshua; Courville, Aaron C.; Vincent, Pascal: Representation Learning: A Review and New Perspectives. IEEE Trans. Pattern Anal. Mach. Intell., 35(8), 2013.

[Bo13] Bordes, Antoine; Usunier, Nicolas; García-Durán, Alberto; Weston, Jason; Yakhnenko, Oksana: Translating Embeddings for Modeling Multi-relational Data. In: NIPS. 2013.

[CBL10] Curran, Clive-Steven; Bröring, Stefanie; Leker, Jens: Anticipating converging industries using publicly available data. Technological Forecasting and Social Change, 77(3), 2010.

[CH74] Caliński, T.; Harabasz, J: A dendrite method for cluster analysis. Communications in Statistics, 3(1), 1974.

[CZS18] Chen, Jianfei; Zhu, Jun; Song, Le: Stochastic Training of Graph Convolutional Networks with Variance Reduction. In: ICML 2018. 2018.

[DBV16] Defferrard, Michaël; Bresson, Xavier; Vandergheynst, Pierre: Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. In: NIPS. 2016.

[De90] Deerwester, Scott; Dumais, Susan T; Furnas, George W; Landauer, Thomas K; Harshman, Richard: Indexing by latent semantic analysis. Journal of the American society for information science, 41(6), 1990.

[DKS12] Duvvuru, Arjun; Kamarthi, Sagar; Sultornsanee, Sivarit: Undercovering research trends: Network analysis of keywords in scholarly articles. In: 2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE). IEEE, 2012.

[ERG19] Eger, S.; Rücklé, A.; Gurevych, I.: Pitfalls in the Evaluation of Sentence Embeddings. arXiv e-prints, June 2019.

[Ge17] Galke, Lukas; et al.: Using Titles vs. Full-text as Source for Automated Semantic Document Annotation. In: K-CAP. ACM, 2017.

[GL16] Grover, Aditya; Leskovec, Jure: node2vec: Scalable feature learning for networks. In: Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2016.

[GVS19] Galke, Lukas; Vagliano, Iacopo; Scherp, Ansgar: Can Graph Neural Networks Go "Online"? An Analysis of Pretraining and Inference. In: Representation Learning on Graphs and Manifolds, ICLR Workshop. 2019.

[HA85] Hubert, Lawrence; Arabie, Phipps: Comparing partitions. Journal of Classification, 2(1), Dec 1985.

[He09] He, Qi; Chen, Bi; Pei, Jian; Qiu, Baojun; Mitra, Prasenjit; Giles, C. Lee: Detecting topic evolution in scientific literature: how can citations help? In: CIKM. ACM, 2009.

[Hu18] Huang, Wen-bing; Zhang, Tong; Rong, Yu; Huang, Junzhou: Adaptive Sampling Towards Fast Graph Representation Learning. In: NeurIPS. 2018.

[HYL17] Hamilton, William L.; Ying, Zhitao; Leskovec, Jure: Inductive Representation Learning on Large Graphs. In: NIPS. 2017.

[Je16] Jeong, Dae-hyun; Cho, Keuntae; Park, Sangyong; Hong, Soon-ki: Effects of knowledge diffusion on international joint research and science convergence: Multiple case studies in the fields of lithium-ion battery, fuel cell and wind power. Technological Forecasting and Social Change, 108, 2016.

[JLC18] Jeong, Daehyun; Lee, Kyuhong; Cho, Keuntae: Relationships among international joint research, knowledge diffusion, and science convergence: the case of secondary batteries and fuel cells. Asian Journal of Technology Innovation, 26(2), 2018.

[KB14] Kingma, Diederik P.; Ba, Jimmy: Adam: A Method for Stochastic Optimization. CoRR, abs/1412.6980, 2014.

[KW16a] Kipf, Thomas N.; Welling, Max: Semi-Supervised Classification with Graph Convolutional Networks. CoRR, abs/1609.02907, 2016. Published at ICLR 2017.

[KW16b] Kipf, Thomas N.; Welling, Max: Variational Graph Auto-Encoders. CoRR, abs/1611.07308, 2016.

[MGS18] Mai, Florian; Galke, Lukas; Scherp, Ansgar: Using Deep Learning for Title-Based Semantic Subject Indexing to Reach Competitive Performance to Full-Text. In: JCDL. ACM, 2018.

[Mi13] Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Gregory S.; Dean, Jeffrey: Distributed Representations of Words and Phrases and their Compositionality. In: NIPS. 2013.

[MRS08] Manning, Christopher D.; Raghavan, Prabhakar; Schütze, Hinrich: Introduction to information retrieval. Cambridge University Press, 2008.

[Ni14] Nitish, Srivastava; Hinton, Geoffrey E.; Krizhevsky, Alex; Sutskever, Ilya; Salakhutdinov, Ruslan: Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 2014.

[NMF17] Niemann, Helen; Moehrle, Martin G.; Frischkorn, Jonas: Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application. Technological Forecasting and Social Change, 115, 2017.

[PAS14] Perozzi, Bryan; Al-Rfou, Rami; Skiena, Steven: DeepWalk: online learning of social representations. In: KDD. ACM, 2014.

[PHW12] Phelps, Corey; Heidl, Ralph; Wadhwa, Anu: Knowledge, Networks, and Knowledge Networks: A Review and Research Agenda. Journal of Management, 38(4), 2012.

[RH07] Rosenberg, Andrew; Hirschberg, Julia: V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. In: EMNLP-CoNLL. ACL, 2007.

[Ro87] Rousseeuw, Peter J.: Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. Journal of Computational and Applied Mathematics, 20, 1987.

[SB88] Salton, Gerard; Buckley, Christopher: Term-weighting approaches in automatic text retrieval. Information processing & management, 24(5), 1988.

[So13] Socher, Richard; Chen, Danqi; Manning, Christopher D; Ng, Andrew: Reasoning With Neural Tensor Networks for Knowledge Base Completion. In: Advances in Neural Information Processing Systems 26. Curran Associates, Inc., 2013.

[SOM17] Salatino, Angelo Antonio; Osborne, Francesco; Motta, Enrico: How are topics born? Understanding the research dynamics preceding the emergence of new areas. PeerJ Computer Science, 3, 2017.

[SOM18] Salatino, Angelo Antonio; Osborne, Francesco; Motta, Enrico: AUGUR: Forecasting the Emergence of New Research Topics. In: JCDL. ACM, 2018.

[Ts09] Tseng, Yuen-Hsien; Lin, Yu-I; Lee, Yi-Yang; Hung, Wen-Chi; Lee, Chun-Hsiang: A comparison of methods for detecting hot topics. Scientometrics, 81(1), 2009.

[URU10] Upham, S. Phineas; Rosenkopf, Lori; Ungar, Lyle H.: Innovating knowledge communities. Scientometrics, 83(2), 2010.

[Ve18] Veličković, Petar; Cucurull, Guillem; Casanova, Arantxa; Romero, Adriana; Liò, Pietro; Bengio, Yoshua: Graph Attention Networks. International Conference on Learning Representations, 2018.

[WJU07] Wuchty, Stefan; Jones, Benjamin F.; Uzzi, Brian: The Increasing Dominance of Teams in Production of Knowledge. Science, 316(5827), 2007.

[Wu19] Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P. S.: A Comprehensive Survey on Graph Neural Networks. arXiv e-prints, January 2019.

[WVC16] Wu, Yan; Venkatramanan, Srinivasan; Chiu, Dah Ming: Research collaboration and topic trends in Computer Science based on top active authors. PeerJ Computer Science, 2, 2016.

[Ya14] Yang, Bishan; Yih, Wen-tau; He, Xiaodong; Gao, Jianfeng; Deng, Li: Embedding Entities and Relations for Learning and Inference in Knowledge Bases. CoRR, abs/1412.6575, 2014.

[YCS16] Yang, Zhilin; Cohen, William W.; Salakhutdinov, Ruslan: Revisiting Semi-Supervised Learning with Graph Embeddings. In: ICML. volume 48 of JMLR Workshop and Conference Proceedings. JMLR.org, 2016.

Automatisierte Analyse Radikaler Inhalte im Internet

Inna Vogel[1], Roey Regev[1], Martin Steinebach[1]

Abstract: Rassismus, Antisemitismus, Sexismus und andere Diskriminierungs- und Radikalisierungsformen zeigen sich auf unterschiedliche Arten im Internet. Es kann als Satire verpackt sein oder als menschenverachtende Parolen. Sogenannte Hassrede ist für die Kommunikationskultur ein Problem, dem die betroffenen Personen oder Personengruppen ausgesetzt sind. Zwar gibt es den Volksverhetzungsparagraphen (§ 130 StGB), Hassrede liegt allerdings nicht selten außerhalb des justiziablen Bereichs. Dennoch sind hasserfüllte Aussagen problematisch, da sie mit falschen Fakten Gruppierungen radikalisieren und Betroffene in ihrer Würde verletzen. 2017 stellte die Bundesregierung das Netzwerkdurchsetzungsgesetz vor, welches die sozialen Netzwerke dazu zwingt, Hassrede konsequent zu entfernen. Ohne eine automatisierte Erkennung ist dieses aber nur schwer möglich. In unserer Arbeit stellen wir einen Ansatz vor, wie solche Inhalte mithilfe des maschinellen Lernens erkannt werden können. Hierfür werden zunächst die Begriffe Radikalisierung und Hate Speech sprachlich eingeordnet. In diesem Zusammenhang wird darauf eingegangen wie Textdaten bereinigt und strukturiert werden. Anschließend wird der k-Nearest-Neighbor-Algorithmus eingesetzt, um Hate Speech in Tweets zu erkennen und zu klassifizieren. Mit unserem Vorgehen konnten wir einen Genauigkeitswert von 0,82 (Accuracy) erreichen - dieser zeigt die Effektivität des KNN-Klassifikationsansatzes.

Keywords: Hassrede, Hate Speech, Soziale Netzwerke, NLP, KNN-Algorithmus, Twitter

1 Einleitung

Die Debattenkultur in sozialen Netzwerken ist nicht selten beleidigend, verletzend, aggressiv oder aber auch hasserfüllt und bedrohlich. Die Sprache radikalisiert sich immer mehr und der verbale Kulturkampf eskaliert seit Jahren zunehmend. Die Feindbilder sind altbekannt: Juden, Linke, Schwarze, Muslime, Homosexuelle, Feministinnen oder Flüchtlinge. Doch Sprache ist nicht nur ein Narrativ. Sie deutet Gegebenheiten und bereitet auf das Handeln vor. Die Diskussionen im Netz dienen nicht nur als Ort des Austausches, sondern auch als Ort konkreter Verabredungen und Planungen von Aktionen, wie aus den folgenden Beispielen abgeleitet werden kann.

- *„So kennt man die Musels. Ist halt ein hinterhältiges Pack. Spielen wir Bimbos versenken ...“*
- *„Diese drecks Bild, wieso steht das nicht auf der 1 Seite-Mauelkorb von Merkel-Leute kauft keine Bild mehr, die veraschen uns so und so nur. Afrikaner...Nigger-schlagt ihn die Köpfe ab, widerliches Pack [...]“*
- *„Lasst uns diese abartigen Asylis entfernen, die haben hier nichts zu suchen“* (Quelle: Twitter)[2]

[1] Fraunhofer-Institut für sichere Informationstechnologie SIT, Rheinstr. 75, 64295 Darmstadt, Germany, {inna.vogel,roey.regev,martin.steinebach}@sit.fraunhofer.de

[2] Die Beispiele in dieser Arbeit illustrieren die Schwere des Problems der Hassrede. Sie stammen aus sozialen Netzwerken und spiegeln in keinster Weise die Meinung der Autoren wider.

doi:10.18420/inf2019_27

„Hassrede“ ist kein juristisch abgegrenzter Begriff, da die unzulässige Meinungsäußerung nicht aus dem jeweiligen Kontext gelöst werden kann. Der Kontext ist meist von der nationalstaatlichen Ordnung geprägt. Als juristischer Ausgangspunkt dient in Deutschland die freie- sowie die unzulässige Meinungsäußerung. Das Grundgesetz stellt in Artikel 5 Abs. 1 Satz 1 fest, dass jeder Mensch das Recht hat, „seine Meinung in Wort, Schrift und Bild frei zu äußern“. Die freie Meinungsäußerung endet jedoch im Artikel 5 Absatz zwei, wenn die persönliche Ehre verletzt wird. Verboten werden kann beispielsweise die Schmähkritik, die stets auf den Kontext ankommt und deshalb eine stets auf den Einzelfall zu prüfende Frage bleibt. Bei Volksverhetzung im Internet droht eine Freiheitsstrafe von bis zu fünf Jahren[3]. Jenseits der schwierigen Rechtslage ist es zudem aufgrund der schieren Masse problematischer Beiträge und Verbreitungsgeschwindigkeit eine Herausforderung Rassismus, Fremdenfeindlichkeit, Antisemitismus oder andere Formen von Intoleranz in ihren Facetten und Dimensionen in den Kommentarspalten des Internets zu identifizieren und einzudämmen.

Anfang 2017 trat das Netzwerkdurchsetzungsgesetz[4] in Kraft. Soziale Netzwerke, darunter Twitter, Facebook und YouTube sind seitdem verpflichtet, „rechtswidrige Inhalte“ innerhalb von 24 Stunden nach Eingang einer Beschwerde zu entfernen oder zu sperren, sonst drohen Bußgelder. Da die Unternehmen hohe Millionen-Strafen fürchten, werden potentiell auch nicht strafbare oder nicht rechtswidrige Inhalte gelöscht[5], was die Meinungsfreiheit der Nutzer einschränkt.

Eine geeignete Lösung hierfür wäre eine Automatisierung oder zumindest eine signifikante semi-automatische Unterstützung, um erfolgreiche Strategien zur Bekämpfung zu entwickeln. In unserer Arbeit stellen wir einen Ansatz vor, wie mithilfe von maschinellen Lernverfahren problematische Beiträge in sozialen Netzwerken automatisch erkannt werden können. Unsere Arbeit ist wie folgt strukturiert: Zunächst werden die Begriffe „Radikalismus“ und „Hassrede“ definiert und voneinander abgegrenzt. In Kapitel 4 stellen wir den verwendeten Textkorpus vor und gehen darauf ein, wie die Twittertexte im Rahmen einer Vorverarbeitung strukturiert und standardisiert werden. Um Hasskommentare automatisch zu klassifizieren, wurde der KNN-Klassifikator (KNN; engl. „k-Nearest-Neighbor“) verwendet. Im Rahmnen unserer Analyse wurden zwei unterschiedliche Frameworks verwendet, deren Unterschied maßgeblich in der Art der Datenrepräsentation liegt. Die vorliegende Arbeit schließt mit einer Zusammenfassung der Ergebnisse und einem Resümee ab.

2 Stand der Forschung

Eine zentrale Herausforderung für das automatische Erkennen von Hasskommentaren in sozialen Medien ist die Trennung zwischen Hate Speech und „nur“ beleidigender Sprache. Davidson et al. [Da17a] haben mithilfe eines Lexikons für beleidigende Sprache Tweets gesammelt und diese anschließend anhand der drei folgenden Kategorien händisch klassifiziert: Hassrede, beleidigende Sprache und neutrale Tweets. Anhand dieses Diktionärs wurde ein Multiklassen-Klassifikator trainiert, um (neue) Tweets anhand dieser drei verschiedenen Kategorien einzuordnen. Eine Analyse der Vorhersagen und Fehler zeigt, dass rassistische und homophobe Tweets eher als Hassrede klassifiziert werden. Sexistische Tweets dagegen werden bevorzugt als beleidigend eingestuft.

Burnap und Williams [BW14] sammelten in den ersten zwei Wochen nach dem Mord an Lee Rigby rund 450.000 Tweets, welche im Zusammenhang mit dem Mord gepostet wurden. Der 25-jährige britische Soldat wurde im Jahr 2013 auf offener Straße in London von zwei mutmaßlichen Islamisten

[3] Strafgesetzbuch (StGB) § 130 Volksverhetzung: https://www.gesetze-im-internet.de/stgb/__130.html

[4] Netzwerkdurchsetzungsgesetz - NetzDG vom 1 September 2017 (BGBl. I p. 3352)

[5] „Soziale Netzwerke löschen tausende Beiträge“: https://www.mdr.de/nachrichten/politik/inland/bilanz-netzdg-twitter-facebook-daten-beschwerden-100.html (29.04.2019)

ermordet. Die Autoren haben unterschiedliche maschinelle Lernverfahren trainiert, um Hasstweets zu klassifizieren, welche sich gegen eine bestimmte Rasse, Ethnie oder Religion richten. Mit ihrem Ansatz konnten sie einen Genauigkeitswert von 0,95 ($F_1 - Score$) erreichen. Der Klassifikator soll Politiker und Entscheidungsträger dabei unterstützen die öffentliche Reaktion auf großräumige emotionale Ereignisse besser einschätzen zu können.

Del Vigna et al. [De17] verwendeten für ihren Ansatz öffentliche Kommentare von italienischen öffentlichen Facebook-Profilen. Nachdem die Hasskommentare von bis zu fünf verschiedenen Annotatoren binär klassifiziert wurden (Hate Speech und neutrale Posts), wurden morpho-syntaktische Merkmale, Sentiment-Polarität und Word Embeddings als Features verwendet, um Hate Speech in Facebook-Kommentaren zu erkennen. Das Trainieren einer Support Vector Machine (SVM) und eines Recurrent Neural Networks (Long Short Term Memory - LSTM) haben die Effektivität der beiden untersuchten Klassifikationsansätze gezeigt.

3 Terminologie

Hassrede (engl. „Hate Speech“)[6] und Radikalisierung sind nicht nur sprachwissenschaftliche, sondern auch politische Phänomene mit Bezügen zu juristischen Tatbeständen. Um radikale Inhalte automatisch zu erkennen, bedarf es eines Algorithmus, der in der Lage ist, Inhalte zu klassifizieren, welche zum Hass gegen Teile der Bevölkerung aufstacheln, zu Gewaltmaßnahmen gegen sie auffordern oder aber die Menschenwürde anderer dadurch angreifen, dass Teile der Bevölkerung beschimpft, böswillig verächtlich gemacht oder verleumdet werden (StGB, §130(1)). Und obwohl Begriffe wie „Radikalismus“, „Extremismus“ und „Hassrede“ im Alltag sowie in den Medien allgegenwärtig sind, ist keineswegs klar, was diese im Einzelnen genau bedeuten. Nicht selten erfolgt ihre Verwendung synonym, da sich ihre Abgrenzung selbst für Wissenschaftler schwierig gestaltet. Um ein gemeinsames Verständnis für die Konzepte zu schaffen, werden im Folgenden die Begriffe kurz definiert und zueinander ins Verhältnis gesetzt.

3.1 Radikalisismus

Radikalismus ist im sozialen und politischen Rahmen eine Geisteshaltung, die eine Änderung von etwas Bestehendem anstrebt [Ne13]. Dies kann beispielsweise das bestehende soziale oder politische System sein. McCauley und Moskalenko [MM08] beschreiben Radikalismus als die Veränderung von Überzeugungen, Gefühlen und Verhaltensmustern. Gruppenkonflikte werden im Zuge dessen zunehmend gerechtfertigt und Opfer zur Verteidigung der Gruppe gefordert. Allerdings muss die Radikalisierung nicht zwingend durch Gewalt gekennzeichnet sein, sondern kann sich in gewaltfreien Widerstands- und Auseinandersetzungsformen manifestieren (z.B. Proteste oder Boykott- und Streikaktionen). Gemäß Wiktorowiczs [Wi05] ist die Radikalisierung ein gradueller und sukzessiver Prozess. Die Person oder Personengruppen passen sich ständig an Normen, Ideologien und Sitten an, die von dem normativen Status quo abweichen. Im Laufe des Radikalisierungsprozesses kommt es zu passiven sowie aktiven Interaktionen eines Individuums in extremistischen Milieus. Dabei werden Handlungen und Ideen befürwortet, die den gängigen Werten der Gesellschaft entgegenstehen [Ne13]. Die Ideologie ist anti-demokratisch, lehnt das bestehende System ab und hat das Ziel dieses mit etwas Neuem zu ersetzen.

Untersuchungen des Bundeskriminalamtes [Bu15] zur Folge steigt die Bedeutung der Radikalisierung in sozialen Medien stetig an, da das Internet extremistischen Gruppen eine (große) Bandbreite an

[6] Die Begriffe Hate Speech, oder deutsch Hassrede, werden nachfolgend synonym verwendet.

Plattformen für den Informationsaustausch und Meinungsäußerung bietet. Zudem erleichtern diese auch die Verbreitung von (Online) Hate Speech und extremistischer Propaganda.

3.2 Hate Speech

Im europäischen Zusammenhang wird Hate Speech zusammengefasst als:

> *„Jegliche Ausdrucksformen, welche Rassenhass, Fremdenfeindlichkeit, Antisemitismus oder andere Formen von Hass, die auf Intoleranz gründen, propagieren, dazu anstiften, sie fördern oder rechtfertigen, einschließlich der Intoleranz, die sich in Form eines aggressiven Nationalismus und Ethnozentrismus, einer Diskriminierung und Feindseligkeit gegenüber Minderheiten und Einwanderern [...] ausdrückt“* (Ministerkomitee des Europarats, Empfehlung R (97) 20, 30.10.1997.

Hate Speech kann sich gegen Hautfarbe, Nationalität, Herkunft, Religion, Geschlecht, sexuelle Orientierung, sozialen Status, Gesundheit, Aussehen, oder eine Kombination davon richten. Die Liste ist keineswegs vollständig, da im Prinzip jede Eigenschaft eines Individuums zum Gegenstand von Hass werden kann [Me13]. Hassäußerungen können unterschiedliche Formen annehmen, sodass es selbst für Menschen nicht immer einfach ist, diese zu entdecken. Sie können direkt oder indirekt geäußert werden. Eine direkte Abwertung von Einwanderern wären beispielsweise folgende Formulierungen: *„Dreckspack entsorgen“*, *„RAUS mit dem PACK“* (Quelle: Twitter). Eine indirekte Herabsetzung wäre beispielsweise: *„Morde, Vergewaltigungen, Messerstechereien...Das ist multi Kulti“* (Quelle: Twitter). Das Gefährliche an Hate Speech ist nicht nur die Verbreitung von verbalen Hassaussagen, sondern oftmals auch die Motivation zum gelebten Gewaltexzess. Hassrede kann zu Übergriffen und Ermordungen an Menschen aufgrund ihrer Hautfarbe, ihrer Religion, ihrer Geschlechtsidentität oder ihrer Sexualität anstiften.

4 Erstellung und Vorverarbeitung der Textdaten

Soziale Netzwerke bieten extremistischen Gruppierungen eine große Auswahl an Plattformen für die Verbreitung von Hassrede. Nicht jeder, der eine Ideologie oder Idee im Netz teilt, ist bereit den Weg der Radikalisierung bis hin zur Gewaltanwendung zu gehen. Allerdings verleitet die vermeintliche Anonymität des Netzes häufig zu sprachlicher Verrohung und zum Verzicht auf Respekt gegenüber Mitmenschen. Sprache wird dabei benutzt, um Ideen und Ideologien zu teilen, bis hin zur Aufforderung von Gewaltanwendung, beispielsweise gegen Einwanderer, Andersgläubige oder andere Minderheiten. Die folgenden Auszüge aus sozialen Medien sollen beispielhaft zeigen wie die Sprache dazu verwendet wird, um Menschen herabzusetzen oder zu verunglimpfen. Das erste Beispiel macht die gefühlte Ungerechtigkeit eines Individuums deutlich sowie den steigenden Rassismus und Ablehnung gegen die Regierung. Das zweite Beispiel offenbart die Vertretung von islamistischer Ideologie. Es wird auf den Gottesstaat referiert, in dem terroristische Gewalt ein Mittel gegen „Ungläubige“ und sogenannte korrupte Regime ist.

> „Politiker bekommen Personenschutz und wer schützt unsere Bevölkerung vor diesen Verbrechern und Abschaum? Das ja den GROSSEN nichts passiert. Haben Angst um ihr bisschen Leben. Was ist mit der Angst die diese alte Frau jetzt hat? Die sollte Schmerzensgeld von der Bundesregierung einklagen die dieses Unrat in unser Land holt.....“ (Quelle: Facebook)

> „DRECKS UNGLÄUBIGE WESTLERABSCHAUM ZIONISTENSCHWEINE PACKT EURE MÄRCHENIDEOLOGIEN UND EUER GRUNDGESETZ HURENSOHN EIN UND VERPISST EUCH BEVOR IHR DER ENDLÖSUNG ZUGEFÜHRT WERDET ; ISIS HÖRT MIT UND DAS SCHARFE SCHWERT REICH FÜR EUCH ZIONISTEN ALLE AUS!!!!! VERPISST EUCH SELBST BEVOR IHR WIE ISRAEL UND DIE RUSSEN SERBEN AUSGERTOTTET WERDET!!!“ (Quelle: YouTube)

Vom „Effekt der sprachlichen Identifikation“ [Di80] wird in der Literatur gesprochen, wenn die Sprache dazu verwendet wird, um sich mit einer Gruppe zu identifizieren [Di80]. Der Effekt besagt, dass innerhalb einer Gruppe dieselbe Sprache gesprochen wird. Dadurch ist es möglich, Hassrede und sich aufbauende Radikalisierung mithilfe von Werkzeugen der linguistischen IT-Forensik und des maschinellen Lernens zu erkennen, bevor es zur Gewaltausübung kommt.

4.1 Korpus

Um Hate Speech automatisch zu identifizieren, haben wir einen englischen Textkorpus mit 25.296 manuell klassifizierten Twitternachrichten[7] [Da17b] verwendet. Twittertexte zeichnen sich dadurch aus, dass jede Meldung eine Maximallänge von 280 Zeichen hat. Die Texte im Korpus wurden von sechs Annotatoren händisch 3 Klassen zugeordnet.

- Klasse 0 = Hate Speech 18.892 Nachrichten
- Klasse 1 = Beleidigende Sprache 1.694 Nachrichten
- Klasse 2 = Neutrale Tweets 1.200 Nachrichten

Der Datenkorpus zeigt, dass die Klassenverteilung ungleich bzw. „unbalanciert“ ist. Es sind deutlich mehr Hate Speech-Tweets im Korpus vorhanden, als neutrale oder beleidigende Nachrichten. Es existieren unterschiedliche Methoden, um dem Problem des Ungleichgewichts im Datensatz entgegenzuwirken. Sampling-basierte Methoden sind in der Datenanalyse Techniken, mit denen die Verteilung der Daten in den Klassen angepasst werden kann. Hierbei unterscheidet man zwischen Undersampling, Oversampling und Hybrid-Verfahren. Hybridverfahren wenden eine Mischform zwischen Under- und Oversampling an, um die Häufigkeit der Daten innerhalb der Klassen anzugleichen.

Beim Undersampling werden Elemente aus der größten Klasse eliminiert, um die Häufigkeitsverteilung der Daten pro Klasse auszugleichen. Das hat allerdings zur Folge, dass weniger Daten der jeweiligen Klasse zum Trainieren zur Verfügung stehen. Beim Oversampling wird die kleinere Klasse synthetisch vervielfältigt, indem beispielsweise zufällig ausgewählte Textdokumente vervielfacht werden. Der Vorteil des Oversamplings besteht darin, dass keine Informationen aus dem Trainingsset verloren gehen, da die Daten sowohl aus der Minderheits- als auch aus der Mehrheitsklasse erhalten bleiben. Der Nachteil ist jedoch, dass der Korpus nicht mit neuen Daten angereichert wird, sondern dieselben Daten lediglich reproduziert werden und dadurch die Gefahr des Overfittings erhöht wird.

Da im verwendeten Korpus Hate Speech über- und neutrale Tweets unterrepräsentiert sind, haben wir uns zunächst für das Oversampling entschieden, um die gleiche Häufigkeitsverteilung der Daten in den Klassen zu erzielen. Statt allerdings synthetisch den Korpus zu erweitern, haben wir die Klasse mit neutralen Tweets erweitert, indem wir den Korpus mit Twittertexten angereichert haben,

[7] Hate Speech Korpus: `https://github.com/t-davidson/hate-speech-and-offensive-language`

die nicht einer einzigen speziellen Domäne zugeordnet werden können. Hierfür wurde der Korpus für Sentimentanalyse von Go et al. [GBH09] verwendet. Der Korpus umfasst rund 1,6 Mio. Tweets sieben unterschiedlicher Domänen (z.B. „Company“, „Event“, „Location“, „Movie“, „Person“ etc.). Dieser Korpus wurde ausgewählt, da die Autoren darauf hinweisen, dass viele Tweets keine Stimmung beinhalten. Vor der Anwendung der Tweets haben wir eine manuelle Überprüfung durchgeführt, um sicherzugehen, dass die Daten nicht aufgrund eines bestimmten Vokabulars oder Topics verzerrt sind. Die Tweets wurden nach dem Zufallsprinzip aus dem Datensatz extrahiert. Es wurde zudem darauf geachtet, dass die Nachrichten in etwa die gleiche Länge und dieselben Textspezifika aufweisen. Twitternachrichten sind speziell, da sie maximal 280 Zeichen lang sind. Zudem werden Hashtags (#), User-Mentions (z.B. @John) sowie Emojis und Emoticons verwendet. Meistens wenden User Umgangssprache bzw. Alltagssprache beim Schreiben an. Der Unterschied zur Fachsprache besteht darin, dass die Alltagssprache, die im täglichen Umgang benutzt wird, keinem bestimmten Soziolekt entspricht wie etwa die Fachsprache (z.B. Wissenschafts- oder Medizinsprache).

Die Domänenunabhägigkeit der Gegenklasse ist wichtig, da sonst der Algorithmus Muster eines bestimmten Themas oder Schreibstils klassifizieren würde[8]. In der folgenden Tabelle sind Beispiele der klassifizierten Hasstweets und neutralen Tweets aus dem verwendeten englischen Datenset aufgeführt:

Hate Speech Tweets	**Neutrale Tweets**
Why people think gay marriage is okay is beyond me. Sorry I don't want my future son seeing 2 fags walking down the street holding hands	Peel up peel up bring it back up rewind back where I'm from they move Shaq from the line„“ ooooow who tf said that trash!!?
#AZmonsoon lot of rain, too bad it wasn't enough to wash away the teabagger racist white trash in the state. #Tcot #teaparty #azflooding	10 birds your grandkids may never see, thanks to climate change http://t.co/XqmXHkAsWt #Climate
#Dutch people who live outside of #NewYorkCity are all white trash.	@SportsCenter: Eli Manning just threw his NFL-leading 27th interception of the season. Lmao trash

Tab. 1: Beispiele für Hasskommentare und neutrale Tweets (Quelle: Twitter)

Um den Datenkorpus zu balancieren, wurde die neutrale Klasse mit zusätzlichen rund 17.996 Tweets erweitert. Je Klasse wurden folglich insgesamt rund 19.200 Tweets verwendet[9]. Die Klasse mit beleidigenden Tweets wurde nach einigen Testversuchen verworfen, da nicht genügend Trainingsmaterial zur Verfügung steht. Allerdings haben wir Tweets, die von mindestens zwei Annotatoren als Hate Speech gelabelt wurden (in der Gesamtbewertung aber als beleidigende Sprache gelabelt wurden), ebenfalls in unserem Korpus aufgenommen. Es wurden folglich die Klassen Hassrede und neutrale Tweets betrachtet und analysiert.

[8] Wenn als Gegenklasse beispielsweise Reviews zu Fotokameras verwendet werden würden, dann könnte der Algorithmus auf der lexikalischen Ebene Unterschiede finden, z.B. zwischen Kamerafeatures und Beleidigungen

[9] Insgesamt weist unser Korpus 19.196 Hate Speech-Kommentare, sowie dieselbe Anzahl an neutrale Tweets, auf

4.2 Datenbereinigung

Um Texte zu strukturieren und zu standardisieren, müssen die Daten einem Preprocessing, d.h. einer Vorverarbeitung, unterzogen werden. Folgende Schritte wurden unternommen:

- Entfernung aller Tweets, die aus weniger als drei Tokens bestehen sowie aller Markup-Tags z.B. <*City*>London<*/City*>
- Entfernung aller User-Mentions (z.B. @John), da diese nicht bedeutungstragend sind
- Kleinschreibung aller Wörter, um das Vokabular zu normalisieren
- Entfernung von Stoppwörtern wie z.B. Artikel („the“, „that“, „a“, „an“), Konjunktionen („and“, „or“, „but“, „because“ etc.) und häufig gebrauchte Präpositionen (z.B. „at“, „in“, „on“ etc.) sowie die Negation „not“. Häufig tragen diese Wortgruppen weniger zum Inhalt bei als die Wortklassen Nomen, Verben, Adjektive und Adverbien
- Entfernung von Sonderzeichen (z.B. <> ()[]{} ∗ +/})
- Negative Smileys und Emojis werden durch CLDR (Common Locale Data Repository) ersetzt, d.h. durch Kurzzeichennamen oder Schlüsselwörter, um diese als Feature nutzen zu können. Die restlichen Smileys werden entfernt. Die Annahme ist, dass negative Smileys und Emojis verwendet werden, um Hate Speech zu untermauern bzw. zu verstärken (z.B. „*Why the fuck do niggers act so different when girls are around?:-S*“)
- Normalisierung der Interpunktion und Buchstabenzeichen. Um dem Gesagten mehr Ausdruck zu verleihen, verwenden Social Media User entweder Großbuchstaben oder eine Aneinanderreihung von gleichen Buchstaben. Damit der Computer erkennt, dass „heeeeeeey“ und „hey“ ein und dasselbe Wort ist, wurden Zeichen, die sich mehr als zweimal wiederholen, zu einem Zeichen reduziert. So wird „heeeeeeey“ zu „hey“ und Wörter wie „arriving“ bleiben unverändert. Angemerkt an dieser Stelle sei, dass dieses triviale Vorgehen zu Fehlern führt. Wird das Wort „Hello“ als „Hellllloooo“ geschrieben, wird es fälschlicherweise zu „Helo“ reduziert. Hier kann mit Rechtschreibprüfungsprogrammen entgegengewirkt werden
- Entfernung von URLs. Es gibt Fälle, bei denen die URL bedeutungstragend ist wie z.B. *„you whore belong to www.pornhub.com“*, aber da diese Fälle selten sind, werden alle URLs gelöscht (z.B. *„www.pornhub.com accepts now Verge, amazing!“*)
- Im letzten Vorverarbeitungsschritt erfolgt die Tokenisierung der Texte

4.3 Hashtags

Hashtags (z.B. *#EarlyChristmas* oder *#FreeMoney*) wurden gesondert behandelt, da diese erheblich zum Inhalt beitragen können. Einerseits sind sie bedeutungstragend, andererseits werden teilweise mehrere Wörter zu einem zusammengefasst, was die maschinelle Verarbeitung erschwert. Zunächst haben wir das Hashtag-Zeichen entfernt und geprüft, ob einem Kleinbuchstaben ein Großbuchstabe folgt. Wenn ja, wurden die Wörter an dieser Stelle voneinander getrennt (z.B. wurde aus *”#FreeMoney”* - *”Free”* und *”Money”*). Anschließend wurde geprüft, ob das Wort in einem Wörterbuch vorhanden ist, wenn ja, wurde es behalten (z.B. *#toys, #shots, #Pisces*). Wenn das Wort nicht in einem Wörterbuch zu finden war, wurde dieses in Fragmente aus drei Buchstabenfolgen, Trigramme genannt, zerlegt und mit den Trigrammen einer Liste mit beleidigenden Wörtern verglichen. Um die Ähnlichkeit der Wörter zu vergleichen, wurde der Jaccard-Koeffizient (J) verwendet. Der Koeffizient nimmt die Mengen A

und B, in unserem Fall die Trigramme, und berechnet daraus den Quotienten der Schnittmenge und deren Vereinigungsmenge ($J(A,B) = \frac{|A \cap B|}{|A \cup B|}$). Der errechnete Wert ist stets zwischen Null und Eins. Je höher der Wert, desto ähnlicher sind sich die Mengen. Die Trigramme werden folglich miteinander verglichen und deren Schnittmenge ermittelt. Hierbei haben wir einen Schwellwert von 0,5 festgelegt. Wurde dieser Schwellwert überschritten, wurde dieses Wort ohne Hashtag im Text behalten. Die restlichen Hashtagwörter wurden entfernt. Der Schwellwert wurde mithilfe eines Sensitivitätstests errechnet. Hierfür wurden 5.000 zufällig gewählte Wörter analysiert. Rund 70% der Wörter konnten durch dieses einfache Verfahren dem richtigen Ursprungswort zugeordnet werden (Beispiel: *"ucunt" - "cunt", "niggas" - "nigger", "abitch" - "bitch"*).

5 Klassifizierung von Hassrede in englischen Twitternachrichten

Um Hate Speech automatisch zu klassifizieren, wurde der Nächste-Nachbarn-Klassifikator (KNN; engl. „k-Nearest-Neighbor“) ausgewählt. Bei diesem Klassifikationsverfahren wird die Klassenzugehörigkeit unter Berücksichtigung der ausgewählten k-nächsten-Nachbarn vorgenommen. Die Dateninstanz wird entweder der Klasse zugeordnet, welche am häufigsten unter den k-Nächsten-Nachbarn vertreten ist oder die Klassenzugehörigkeit erfolgt gemäß des nächsten Nachbarns, gemessen an der geringsten Distanz (k=1). Trotz der relativ simplen Funktionsweise gehört der Klassifikator zu den erfolgreichsten maschinellen Lernverfahren [Ko15].

Die Funktionsweise des KNN-Algorithmus ist in Abbildung 1 beschrieben. Es wird ein binäres Klassifikationsproblem mit den Klassen A und B betrachtet. Werden k=3 nächste Nachbarn in Betracht gezogen, erfolgt die Zuordnung des unbekannten Objekts zur Klasse B. Werden k=6 nächste Nachbarn ausgewählt, ähnelt das Objekt laut dem Algorithmus der Klasse A. Hier offenbart sich ein Nachteil des Verfahrens. Wird ein zu „kleines“ k ausgewählt (z.B. k=2), weist der Algorithmus eine hohe Sensitivität gegenüber Ausreißern auf. Wird k zu hoch angesetzt (z.B. k=30), werden zu viele Objekte der Gegenklasse in der Entscheidungsmenge vorkommen und vom Klassifikator „favorisiert“.

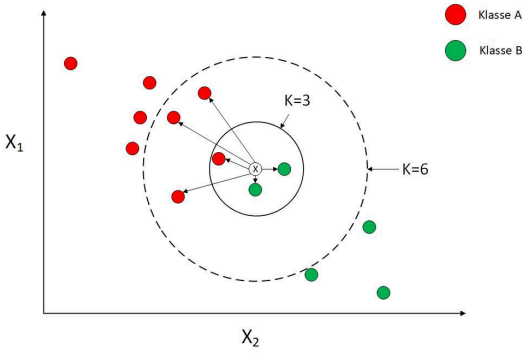


Abb. 1: KNN-Klassifikator mit k=3 und k=6 Nachbarn

5.1 NuPIC KNN-Framework

Der KNN-Klassifikator wurde auf zwei unterschiedliche Weisen verwendet. Das erste Framework wurde mittels „NuPIC“ (Numenta Platform for Intelligent Computing) integriert, einer Pythonbibliothek für den HTM-Lernalgorithmus (HTM - engl. „Hierarchical Temporal Memory“). HTM entstammt aus dem Forschungsgebiet der Neurowissenschaft und wurde von Jeff Hawkins und George Dileep [Ha19] entwickelt. HTM ist eine detaillierte Berechnungstheorie, welche versucht, die strukturellen und algorithmischen Eigenschaften des Neokortex in einem Bayes'schen Netz abzubilden. Das Herzstück von HTM sind zeitbasierte, kontinuierliche Lernalgorithmen, die räumliche und zeitliche Muster speichern und abrufen. NuPIC bietet neben dem HTM-Lernalgorithmus unterschiedliche Klassifikatoren an[10], von denen wir den „KNN Classifier“ verwendet haben.

Als Eingabe erfordert NuPICs KNN eine feste Länge des Eingabevektors, weshalb alle Tweets auf dieselbe Zeichenlänge aufgefüllt wurden. Als Eingabevektor wurde die Maximallänge der Tweets von 280 Zeichen gewählt. Die Eingabedaten müssen bei NuPIC spärlich repräsentiert sein (SDR; engl. „Sparse Distributed Representation“). Für das Encoding wurden die Buchstabenzeichen folglich binär in One-Hot-Vektoren umgewandelt (z.B. [00101010] = [2,4]). Der „SparsePassThroughEncoder“ wurde verwendet, um die One-Hot-Vektoren an den KNN-Klassifikator zu übergeben.

Während der Trainingsphase wird eine Menge der zur Klassifikation ausgewählten k-nächsten Nachbarn betrachtet. Je ähnlicher die Dokumente, desto näher zueinander befinden sich diese im n-dimensionalen Vektorraum, wobei n die Dimension des Eingabevektors ist. Je unterschiedlicher die Dokumente, desto weiter sind diese im Vektorraum voneinander entfernt. In der Testphase wurde abschließend die euklidische Distanz genutzt, um zu eruieren, wie präzise das Verfahren klassifiziert. Andere Distanzmaße wie bspw. die Manhatten-Distanz sind ebenso denkbar. Die Testergebnisse haben eine Accuracy von 0,71 erzielt. Das Entfernen der Stoppwörter resultiert in einer marginalen Verschlechterung.

Der Genauigkeitswert kann weiter gesteigert werden, indem die Wörter im Korpus beispielsweise auf ihre Grundform reduziert werden. Dieses sogenannte „Stemming“ (Grundformenreduktion) ist ein Verfahren, mit dem verschiedene morphologische Varianten eines Wortes auf ihren gemeinsamen Wortstamm (stem) zurückgeführt werden, z.B. „essen“ - „isst“ - „gegessen“ = „ess“. Dieser Ansatz reduziert das Vokabular deutlich und führt bei unterschiedlichen Verfahren zu einer besseren Klassifizierungsperformance.

5.2 Scikit-learn KNN-Framework

Das zweite KNN-Framework wurde mittels „scikit-learn“ integriert. Der Unterschied zum ersten Framework ist hauptsächlich in der Datenrepräsentation begründet. Um die Twitternachrichten zu vektorisieren, wurde das Tf-idf-Maß (engl. „Term Frequency / Inverse Document Frequency“) verwendet. Das Maß wird dazu verwendet, um die Relevanz eines Terms im Dokument im Vergleich zur Dokumentenkollektion zu beurteilen. Zunächst wird die Termfrequenz berechnet, d.h. wie häufig ein Wort im Dokument verwendet wird. Anschließend wird ermittelt in wie vielen Dokumenten eine Kollektion dieses Wortes vorkommt. Die Relevanzhypothese besagt, dass Schlüsselbegriffe, die in einem Dokument (themenbezogen) relativ häufig vorkommen, in der Gesamtheit der Dokumente aber relativ selten, ein guter Indikator für den Inhalt eines Dokumentes sind. Artikel und Konjunktionen

[10] NuPIC API Documentation: `http://nupic.docs.numenta.org/1.0.0/index.html`

kommen dagegen in allen Dokumenten etwa gleich oft vor und sind deshalb weniger relevant für die Bestimmung eines Topics. Diese sind beispielsweise relevant für die Stilerkennung eines Textes.

$$TF(i,j) = \frac{\text{Häufigkeit des Begriffs i im Dokument j}}{\text{Gesamtwörter im Dokument j}}$$

$$IDF(i) = \frac{\text{Gesamtdokumente}}{\text{Dokumentananzahl mit Term i}}$$

$$TF - IDF = TF(i,j) * IDF(i)$$

Eine andere Methode, um Wörter zu repräsentieren, wäre der Einsatz von „Word Embeddings“. Word Embeddings stellen Wörter nicht nur mathematisch dar, sondern repräsentieren darüber hinaus die Bedeutung der Wörter im Kontext von anderen Wörtern. Dabei werden Begriffe anhand ihres Vorkommens im Textkorpus und der sie umgebenden Wörter so in einem multidimensionalen Raum angeordnet, dass Wörter, die im selben Kontext vorkommen, einen ähnlichen Vektor erhalten. Mittlerweile existieren vortrainierte Modelle wie etwa „fastText“ oder „GloVe“. Wenn genügend Daten vorhanden sind, können eigene Word Embeddings trainiert werden. Das folgende Beispiel zeigt Ähnlichkeitsmaße eines eigens trainierten Word-Embeddings-Modells:

- „Asylanten“ = „Asylforderer“ (0,73), „Flüchtlinge“ (0,72), „Krimigranten“ (0,70), „Invasoren“ (0,69), „Migranten“ (0,67)
- „KZ“ = „Räucherhaus“ (0,68), „Vernichtungslager“ (0,65), „Auschwitz“ (0,63)
- „ausrotten“ = „hassen“ (0,61), „abschlachten“ (0,58), „Zionisten“ (0,58), „töten“ (0,57)

Nachdem jedes Wort einen Tf-idf-Wert als Merkmal erhalten hat, wurde der KNN-Algorithmus trainiert. 70% der Rohdaten wurden als Trainings- und 30% als Testset verwendet. Auch hier wurde nach dem Training die euklidische Distanz genutzt, um die Performance des Algorithmus zu eruieren (k=5 Nachbarn). Die Accuracy liegt bei diesem Ansatz bei 0,82. Diese Ergebnisse zeigen die Effektivität der beiden Klassifikationsansätze.

6 Evaluierung

Um Hassrede automatisch zu identifizieren und zu klassifizieren, haben wir einen Textkorpus mit 25.296 manuell klassifizierten Twitternachrichten verwendet. Da die Klassenverteilung ungleich war, haben wir die kleinere Klasse mit domänenunabhängigen Tweets erweitert. Dieser Schritt war nötig, da KNN als Klassifikationsalgorithmus trainiert und evaluiert wurde. Beim KNN-Klassifikator sollte darauf geachtet werden, dass die Daten in den Klassen in etwa gleich verteilt sind, damit das Verfahren die stärker vorkommende Klasse bei der gewählten k-nächsten-Nachbarn Betrachtung nicht „bevorzugt“.

Um den KNN-Klassifikationsalgorithmus zu trainieren und dessen Performance anschließend zu untersuchen, haben wir die Frameworks NuPIC und Scikit-Learn verwendet. Für die Verwendung des NuPIC KNNs wurden als Feature-Vektoren Buchstabenzeichen binär in One-Hot-Vektoren umgewandelt. Bei der zweiten Verwendung mittels Scikit-Learn wurden Tokens nach dem Tf-idf-Maß gewichtet und als Datenrepräsentation eingesetzt. Nach dem Training wurde die euklidische Distanz genutzt, um die Performance des Algorithmus zu quantifizieren.

Die Ergebnisse der beiden Klassifikatoren werden in der in Tabelle 2 dargestellten Konfusionsmatrix zusammen mit den resultierenden Evaluierungsmaßen Accuracy (Acc.), Precison (P), Recall (R) und dem F_1-Score, dargestellt. Die Konfusionsmatrix stellt mit ihren vier möglichen Ausprägungen[11] die Grundlage für die Evaluierung eines Großteils der binären Klassifikationsverfahren dar.

Methode & Features	**Konfusionsmatrix**				**Performanzmaß**			
	TP	**TN**	**FP**	**FN**	**Acc.**	**P**	**R**	**F1**
NuPIC KNN One-Hot-Vekt.	4.875	3.253	2.487	869	0,71	0,66	0,85	0,74
Scikit-learn KNN Tf-idf	4.464	4.943	814	1.293	0,82	0,85	0,84	0,82

Tab. 2: Klassifikationsgüte der KNN-Klassifikatoren

Die besten Klassifikationsergebnisse erzielte der KNN Klassifikator, wenn die Tokens als Tf-idf-Vektoren repräsentiert wurden. Hierbei wurde eine Accuracy von 0,82 erzielt. Wurden die Daten als Buchstabenzeichen binär in One-Hot-Vektoren umgewandelt, wurde eine Klassifikationsgüte von 0,71 erreicht. Beim Vergleich der FP- und FN-Werte zeigt sich, dass beim Scikit-Learn-Verfahren mehr Hate Speech Tweets fälschlicherweise als neutrale Tweets klassifiziert wurden. Das könnte damit zusammenhängen, dass die Alltags- und Umgangssprache in sozialen Netzwerken eine starke Verwendung findet. Hierzu gehört auch der Gebrauch von Slang und Schimpfwörtern. Diese klassifiziert der Algorithmus fälschlicherweise als Hate Speech. Davidson et al. [Da17a] haben in diesem Zusammenhang angemerkt, dass die Wörter *„fag“, „bitch“* oder *„nigga“* sowohl Slang sein können, als auch eine Ausdrucksform von Rassismus, Fremdenfeindlichkeit, Sexismus oder Homophobie. Auch in den neutralen Tweets, die wir für Trainingszwecke verwendet haben, finden sich vermehrt Slang- und Schimpfwörter, die ebenfalls als Form der Intoleranz gegenüber Personen- und Personengruppen eingesetzt werden können (z.B. *„pussy“, „bitch“, „trash“*).

- *„Overdosing on heavy drugs doesn’t sound bad tonight. I do that pussy shit every day.“*
- *„a pissed lad past out. I would lick his dirty soles while he slept.“*
- *„welfare/government aid is claimed by white people. So y’all black slander is trash now.“*
- *„http://t.co/JOOsdSubIR he’s a bitch“*

7 Zusammenfassung und Ausblick

Radikalisierungs- und Diskriminierungsformen zeigen sich auf unterschiedliche Weise in sozialen Netzwerken. Hassrede liegt dabei aber nicht selten außerhalb des justiziablen Bereichs. Dennoch sind Hassaussagen problematisch, da sie beispielsweise mit falschen Fakten Menschen oder Gruppierungen radikalisieren können. Ohne eine automatisierte Erkennung ist deren Eindämmung kaum möglich.

In unserer Arbeit haben wir den KNN-Algorithmus eingesetzt, um Hate Speech in Tweets zu erkennen und zu klassifizieren. Zunächst haben wir die Begriffe Radikalisierung und Hate Speech sprachlich definiert und eingeordnet. Anschließend haben wir unterschiedliche Verfahren vorgestellt, wie Textdaten bereinigt und strukturiert werden können. Da im verwendeten Korpus Hate Speech über- und neutrale Tweets unterrepräsentiert waren, haben wir die Klasse mit neutralen Tweets erweitert, indem wir den Korpus mit Twittertexten angereichert haben, die keiner speziellen Domäne zugeordnet waren. Je Klasse wurden rund 19.200 Tweets verwendet. Die Klasse mit beleidigenden Tweets wurde

[11] True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN)

nach einigen Testversuchen verworfen, da nicht genügend Trainingsmaterial zur Verfügung stand. Für KNN ist es essenziell, dass die Häufigkeit der Datenverteilung in den Klassen in etwa gleich ist.

Der Algorithmus KNN wurden auf zwei unterschiedliche Weisen eingesetzt - mittels des „NuPIC"- und des „scikit-learn"-Frameworks. Die besten Klassifikationsergebnisse hat KNN erzielt, wenn die Tokens als Tf-idf-Vektoren repräsentiert wurden. Hierbei wurde eine Accuracy von 0,82 erzielt. Wurden die Daten als Buchstabenzeichen binär in One-Hot-Vektoren umgewandelt, wurde eine Klassifikationsgüte von 0,71 erreicht.

Die Evaluierung hat ergeben, dass neutrale Tweets eher fälschlicherweise als Hate Speech klassifiziert werden, als umgekehrt. Das hängt mitunter damit zusammen, dass in sozialen Netzwerken Slang- und Schimpfwörter nicht nur als Ausdrucksform von Rassismus, Fremdenfeindlichkeit, Sexismus oder Homophobie verwendet werden. Untersuchungen haben gezeigt, dass rassistische und homophobe Tweets eher als Hassrede klassifiziert werden. Sexistische Tweets dagegen werden bevorzugt als beleidigend eingestuft. Für die zukünftige Forschung ist es deshalb essenziell einen hinreichend großen Korpus mit beleidigenden Texten zu labeln, um Verfahren zu trainieren, die nicht nur zwischen neutralen und Hasskommentaren unterscheiden können, sondern auch den Unterschied zwischen „nur" beleidigender Sprache und Hate Speech erkennen können. Zudem wollen wir mit anderen Features als Eingabevektoren experimentieren, wie beispielsweise mit den bereits erwähnten Word Embeddings oder indem die Wörter auf ihre Grundform reduziert werden.

Um auch radikale Inhalte maschinell klassifizieren zu können, bedarf es einer hinreichenden Menge von Trainingsdaten, welche zuvor von Experten händisch klassifiziert wurden. Hierfür muss allerdings im Vorfeld klar definiert werden, welche radikalen Aussagen bereits auf das Handeln vorbereiten. Die Textmuster werden anschließend für maschinelle Lernverfahren eingesetzt. In diesem Kontext muss klar abgegrenzt sein was noch als freie Meinungsäußerung gilt und was bereits rechtswidrig ist. Denn das Ziel der frühzeitigen Erkennung von Hate Speech und Radikalisierung ist es, Rechtswidrigkeit einzudämmen, nicht das subjektive Recht auf freie Rede sowie freie Meinungsäußerung einzuschränken.

Literaturverzeichnis

[Bu15] Bundeskriminalamt: Analyse der Radikalisierungshintergründe und -verläufe der Personen, die aus islamistischer Motivation aus Deutschland in Richtung Syrien oder Irak ausgereist sind. Wiesbaden, 2015.

[BW14] Burnap, Peter; Williams, Matthew Leighton: Hate speech, machine classification and statistical modelling of information flows on twitter: Interpretation and communication for policy decision making. 2014.

[Da17a] Davidson, Thomas; Warmsley, Dana; Macy, Michael; Weber, Ingmar: Automated hate speech detection and the problem of offensive language. In: Eleventh International AAAI Conference on Web and Social Media. 2017.

[Da17b] Davidson, Thomas; Warmsley, Dana; Macy, Michael; Weber, Ingmar: Automated Hate Speech Detection and the Problem of Offensive Language. In: Proceedings of the 11th International AAAI Conference on Web and Social Media. ICWSM '17, S. 512–515, 2017.

[De17] Del Vigna12, Fabio; Cimino23, Andrea; Dell'Orletta, Felice; Petrocchi, Marinella; Tesconi, Maurizio: Hate me, hate me not: Hate speech detection on Facebook. 2017.

[Di80] Dieckmann, Walther: Sprache in der Politik? Die Rolle der Sprache in der Politik. Carl Hanser VerlG, 1980.

[GBH09] Go, Alec; Bhayani, Richa; Huang, Lei: Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford, 1(12):2009, 2009.

[Ha19] Hawkins, Jeff; Lewis, Marcus; Klukas, Mirko; Purdy, Scott; Ahmad, Subutai: A Framework for Intelligence and Cortical Function Based on Grid Cells in the Neocortex. Frontiers in Neural Circuits, 12:121, 2019.

[Ko15] Koch, Dominik: Verbesserung von Klassifikationsverfahren: Informationsgehalt der k-Nächsten-Nachbarn nutzen. BestMasters. Springer Fachmedien Wiesbaden, 2015.

[Me13] Meibauer, Jörg: Hassrede : von der Sprache zur Politik. In: Hassrede - hate speech : interdisziplinäre Beiträge zu einer aktuellen Diskussion. Gießener Elektronische Bibliothek, Gießen, S. 1–16, 2013.

[MM08] McCauley, Clark; Moskalenko, Sophia: Mechanisms of political radicalization: Pathways toward terrorism. Terrorism and political violence, 20(3):415–433, 2008.

[Ne13] Neumann, Peter: Radikalisierung, Deradikalisierung und Extremismus. Aus Politik und Zeitgeschichte, 63(29-31):3–10, 2013.

[Wi05] Wiktorowicz, Q.: Radical Islam Rising: Muslim Extremism in the West. G - Reference,Information and Interdisciplinary Subjects Series. Rowman & Littlefield, 2005.

Extended Abstracts

The DESQ Framework for Declarative and Scalable Frequent Sequence Mining

Presentation of work originally published in IEEE 16th Intl. Conf. on Data Mining, IEEE 35th Intl. Conf. on Data Engineering, and 2019 ACM Trans. on Database Syst.

Kaustubh Beedkar[1], Rainer Gemulla[2], Alexander Renz-Wieland[3]

Abstract: DESQ is a general-purpose framework for declarative and scalable frequent sequence mining. Applications express their specific sequence mining tasks using a simple yet powerful powerful pattern expression language, and DESQ's computation engine automatically executes the mining task in an efficient and scalable way. In this paper, we give a brief overview of DESQ and its components.

Keywords: Data mining; Frequent sequence mining; Subsequence constraints; Pattern expressions; Distributed sequence mining

Frequent sequence mining (FSM, [Do09]) is a fundamental task in data mining: Given a sequence database, the task is to find interesting sequential patterns that appear frequently in the data. In customer behavior analysis, for example, frequent sequences may correspond to purchase patterns of customers, or to popular navigations paths across a website, and can serve as an input for applications such as recommender systems. The task arises in a wide range of applications, including natural language processing, information extraction, web usage mining, market-basket analysis, and computational biology.

DESQ is a general-purpose framework for FSM that aims to support such a wide range of applications. DESQ features (1) a powerful pattern expression language that allows applications to describe declaratively which patterns are considered interesting, (2) a suite of efficient and scalable mining algorithms that support both sequential and distributed execution, and (3) an easy-to-use API based on Apache Spark. The DESQ framework allows applications to express a wide range of sequence mining problems—including and beyond those considered in prior literature—in a unified way. This unified treatment improves the usability of pattern mining in practice: Data scientists only need to familiarize themselves with one framework and, perhaps more importantly, do not need to develop customized mining algorithms for a particular application. Likewise, a unified treatment allows researchers to study jointly many variants of FSM, instead of each one individually.

Consider for example the task of mining frequent relational phrases between entities from large text corpora; e.g., the phrase *make a deal with* may be frequent between persons and/or organizations. Such patterns are indicative of relations between entities and arise in natural language processing and information extraction applications. Existing

[1] Technische Universität Berlin, kaustubh.beedkar@tu-berlin.de
[2] Universität Mannheim, rgemulla@uni-mannheim.de
[3] Technische Universität Berlin, alexander.renz-wieland@tu-berlin.de

doi:10.18420/inf2019_28

Tab. 1: Example pattern expressions for some FSM tasks and frequencies in New York Times data (first two blocks) and Amazon review data (last block).

Pattern expression	FSM task	Example patterns (frequency)
(.){1,4}	*n-grams of up to four words*	green tea (337), editor in chief (3275)
(.)[.{0,2}(.)]{1,3}	*Skip n-grams of 2–4 words with gap at most 2*	flight from to (758), son of and of (15896)
ENTITY (VERB$^+$ DET? NOUN$^+$? PREP?) ENTITY	*Relational phrases*	is being advised by (15), has coached (10)
(ENTITY$^\uparrow$ VERB$^+$ NOUN$^+$? PREP? ENTITY$^\uparrow$)	*Typed relational phrases*	ORG headed by ENTITY (275), PER born in LOC (481)
(Book)[.{0,2}(Book)]{1,4}	*Sequences of books*	'A Storm of Swords' 'A Feast for Crows'(153)
DigitalCamera[.{0,3}(.$^\uparrow$)]{1,4}	*Products or types of products purchased after a digital camera*	'Lenses' 'Tripods'(158), 'Batteries' 'SD&SDHC Cards'(149)

FSM algorithms cannot solve such a task since they cannot be tailored to consider only relational phrases (thereby producing many uninteresting—i.e., non-relational—patterns) or to consider context information (thereby producing patterns that do not connect entities). In contrast, this mining task can be expressed in DESQ's pattern expression language as *ENTITY* (*VERB*$^+$ *DET*? *NOUN*$^+$? *PREP*?) *ENTITY*.

DESQ's pattern expression language is based on regular expressions and additionally includes features such as item hierarchies and capture groups. In the above example, item hierarchies allow applications to relate items to each other (e.g., *make* is a *VERB*), and capture groups allow to express what is considered part of the pattern (the relational phrase) and what context (between entities). Table 1 lists some additional examples, in which pattern expressions are used to either concisely describe traditional FSM tasks or to define customized sequence mining tasks.

DESQ includes a number of general-purpose mining algorithms for the wide range of mining tasks that can be expressed using pattern expressions. In particular, DESQ provides efficient algorithms that can operate on a single machine as well as scalable, distributed algorithms. The pattern expression language, the underlying computational framework, and efficient mining algorithms are described in [BG16; BGM19; RBG19].

DESQ is available as an open source library.[4] The library provides a Java API for a single machine setup, and a Scala API for a distributed setup on top of Apache Spark. The API allows applications to perform pattern mining directly on datasets in their native formats.

Literatur

[BG16] Beedkar, K.; Gemulla, R.: DESQ: Frequent Sequence Mining with Subsequence Constraints. In: ICDM. S. 793–798, 2016.

[BGM19] Beedkar, K.; Gemulla, R.; Martens, W.: A Unified Framework for Frequent Sequence Mining with Subsequence Constraints. ACM Trans. Database Syst. 44/3, 11:1–11:42, 2019.

[Do09] Dong, G.: Sequence Data Mining. Springer-Verlag, Berlin, Heidelberg, 2009.

[RBG19] Renz-Wieland, A.; Bertsch, M.; Gemulla, R.: Scalable Frequent Sequence Mining With Flexible Subsequence Constraints. In: ICDE. S. 1490–1501, 2019.

[4] `https://www.uni-mannheim.de/dws/research/resources/desq`

Adversarial Attacks on Graph Neural Networks

Presentation of work originally published in the Proc. of the 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining as well as the International Conference on Learning Representations 2019

Daniel Zügner[1], Amir Akbarnejad[1], Stephan Günnemann[1]

Keywords: deep learning; graph neural networks; adversarial machine learning

Graphs are at the core of many high impact applications ranging from the analysis of social and rating networks (Facebook, Amazon), over gene interaction networks (BioGRID), to interlinked document collections (PubMed, Arxiv). Deep learning models for graphs have advanced the state of the art on many tasks such as node classification or link prediction; they are currently being deployed in production systems, e.g. for content recommendation on social media [Yi18]. Despite their recent success, little is known about their robustness. Yet, in domains where they are likely to be used, e.g. the web, adversaries are common. Can deep learning models for graphs be easily fooled? In [ZAG18, ZG19] we introduce the first studies of adversarial attacks on graph neural networks, aiming to reduce their performance by adding small perturbations to the data. In addition to attacks at test time, we tackle the more challenging class of poisoning/causative attacks, which focus on the training phase of a machine learning model. We generate adversarial perturbations targeting the *node features* and the *graph structure*, thus, taking the dependencies between instances in account. Moreover, we ensure that the perturbations remain *unnoticeable* by preserving important data characteristics. We propose two algorithms: one for *targeted* attacks whose goal is to misclassify a specific target node and one for *global* adversarial attacks aiming to reduce the overall classification performance on test data. The former exploits incremental computations for efficient targeted attacks, and the latter uses meta-gradients to directly tackle the bilevel problem underlying training-time attacks, essentially treating the graph as a hyperparameter to optimize. Our experiments show that small graph perturbations consistently lead to a strong decrease in performance for graph convolutional networks, transfer to unsupervised embeddings, and likewise are successful even when only limited knowledge about the graph is given. Remarkably, the perturbations created by our global attack algorithm can misguide the graph neural networks such that they perform *worse* than a linear classifier that ignores all relational information. Our findings emphasize that further research is needed to improve the robustness of graph neural networks.

[1] Technical University of Munich [zuegnerd,akbarnej,guennemann]@in.tum.de

doi:10.18420/inf2019_29

References

[Yi18] Ying, Rex; He, Ruining; Chen, Kaifeng; Eksombatchai, Pong; Hamilton, William L; Leskovec, Jure: Graph convolutional neural networks for web-scale recommender systems. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. ACM, pp. 974–983, 2018.

[ZAG18] Zügner, Daniel; Akbarnejad, Amir; Günnemann, Stephan: Adversarial attacks on neural networks for graph data (*Best Paper Award*). In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 2847–2856, 2018.

[ZG19] Zügner, Daniel; Günnemann, Stephan: Adversarial Attacks on Graph Neural Networks via Meta Learning. In: International Conference on Learning Representations (ICLR). 2019.

Finding Tiny Clusters in Bipartite Graphs (Extended Abstract)

Presentation of work originally published in the Proceedings of the Thirty-second Annual Conference on Neural Information Processing Systems (NeurIPS 2018) under the title *Bipartite Stochastic Block Models With Tiny Clusters* [Ne18]

Stefan Neumann [1]

Abstract: We study the problem of finding clusters in random bipartite graphs. Applications of this problem include online shops in which one wants to find customers who purchase similar products and groups of products which are frequently bought together. We present a simple two-step algorithm which provably finds *tiny* clusters of size $O(n^{\varepsilon})$, where n is the number of vertices in the graph and $\varepsilon > 0$; previous algorithms were only able to identify medium-sized clusters consisting of at least $\Omega(\sqrt{n})$ vertices. We practically evaluate the algorithm on synthetic and on real-world data; the experiments show that the algorithm can find extremely small clusters even when the graphs are very sparse and the data contains a lot of noise.

Keywords: Biclustering; Bipartite Graphs; Random Graphs; Stochastic Block Models

1 Introduction

Finding clusters in bipartite graphs is a fundamental problem and has many applications. In practice, the two sides of the bipartite graph usually correspond to objects from different domains and an edge corresponds to an interaction between the objects. For example, in an online shop setting, the bipartite graph consists of customers (left side of the graph) and products (right side of the graph). An edge indicates that a customer bought a certain product. In this scenario, *customer clusters* consist of customers buying similar items and *product clusters* consist of products which are frequently bought together. Other application domains include paleontology [Fo03], where one wants to find co-occurrences of localities and mammals, and bioinformatics [Er13], where one wants to relate biological samples and gene expression levels.

Note that in many practical scenarios it is important that one can find *tiny* clusters. For example, nowadays online shops sell millions of products, but most product clusters are

[1] Universität Wien, Fakultät für Informatik, Wien, Österreich. stefan.neumann@univie.ac.at. The author gratefully acknowledges the financial support from the Doctoral Programme "Vienna Graduate School on Computational Optimization" which is funded by the Austrian Science Fund (FWF, project no. W1260-N35).

doi:10.18420/inf2019_30

very small compared to the total number of products on sale (e.g., the *Harry Potter* books are often bought together but they are only seven out of more than one million products). Hence, if a clustering algorithm can only detect medium-sized clusters consisting of at least a thousand products, then it is not applicable in this setting.

In this paper, we study this problem under a standard random graph model with a set of planted ground-truth clusters and we propose an algorithm which *provably* recovers extremely small clusters. More formally, we show that the algorithm allows to recover even tiny planted clusters of size $O(n^\varepsilon)$, where n is the number of vertices on the right side of the graph and $\varepsilon > 0$. Previous methods [Xu14, LCX15] could only discover clusters of size $\Omega(\sqrt{n})$. For the formal statement of the random graph model and the results, see [Ne18].

The algorithm consists of a simple two-step procedure: (1) Cluster the vertices on the left side of the graph based on the similarity of their neighborhoods. (2) Infer the right-side clusters based on the previously discovered left clusters using degree-thresholding.

We also implement the algorithm and evaluate it on synthetic and on real-world data. We verify that, in practice, the algorithm can find the small clusters which the theoretical analysis promised. We answer this question affirmatively. On synthetic data, the experiments show that, indeed, the algorithm finds tiny clusters even in the presence of high destructive noise (i.e., when the graphs are sparse and there are few inter-cluster edges). On real-world datasets, the algorithm finds clusters which are interesting and which have natural interpretations; for example, on a dataset consisting of users and books they rated [Zi05], the algorithm finds (among others) one cluster consisting of the *Harry Potter* books by J. K. Rowling and another cluster consisting of books written by John Grisham.

Bibliography

[Er13] Eren, Kemal; Deveci, Mehmet; Küçüktunç, Onur; Çatalyürek, Ümit V.: A comparative analysis of biclustering algorithms for gene expression data. Briefings in Bioinformatics, 14(3):279–292, 2013.

[Fo03] Fortelius, M. (coordinator): , New and Old Worlds Database of Fossil Mammals (NOW). Online. `http://www.helsinki.fi/science/now/`, 2003. Accessed: 2015-09-23.

[LCX15] Lim, Shiau Hong; Chen, Yudong; Xu, Huan: A Convex Optimization Framework for Bi-Clustering. In: ICML. pp. 1679–1688, 2015.

[Ne18] Neumann, Stefan: Bipartite Stochastic Block Models with Tiny Clusters. In: Thirty-second Conference on Neural Information Processing Systems, NeurIPS 2018. pp. 3871–3881, 2018.

[Xu14] Xu, Jiaming; Wu, Rui; Zhu, Kai; Hajek, Bruce E.; Srikant, R.; Ying, Lei: Jointly clustering rows and columns of binary matrices: algorithms and trade-offs. In: SIGMETRICS. pp. 29–41, 2014.

[Zi05] Ziegler, Cai-Nicolas; McNee, Sean M.; Konstan, Joseph A.; Lausen, Georg: Improving recommendation lists through topic diversification. In: WWW. pp. 22–32, 2005.

DipTransformation: Enhancing the Structure of a Dataset and thereby improving Clustering (Extended Abstract)

Presentation of work originally published in the Proceedings of the 2018 IEEE International Conference on Data Mining (ICDM 2018).

Benjamin Schelling[1] Claudia Plant[1 2]

Keywords: Clustering, Dip-test, Dataset-Transformation

The clustering of a data set depends strongly on the structure it contains. A data set might have a well-defined structure, but this does not necessitate good clustering results. If the structure is hidden in an unfavourable scaling, clustering usually fails. Confronted with a data set one cannot quite cluster, usually this would lead to a new clustering method which is capable of dealing with the new and problematic type of data set, but this is not always necessary. The aim of the DipTransformation is to enhance the data set by re-scaling and transforming its features and thus emphasizing and accentuating its structure. If the structure is sufficiently clear, clustering algorithms - even well-established ones - will perform far better. To the best of our knowledge, there are currently no methods besides DipTransformation that have the goal of enhancing structure.

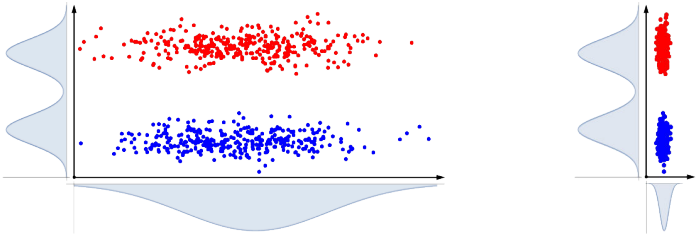
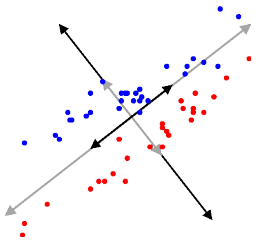


Fig. 1: A simple, synthetic data set before (a) and after (b) scaling it with its dip values. The dip value are used to tell us how much structure a feature contains and how relevant it is for clustering.

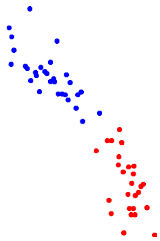
DipTransformation makes it possible to compensate for the unfortunate scaling of the features with the help of the Dip test [1]. The Dip test measures the amount of structure in a feature. Take a look at Fig. 1. It is a very simple data set, consisting of two Gaussian

[1] Faculty of Computer Science, University of Vienna, Vienna, Austria
[2] ds:UniVie, University of Vienna, Vienna, Austria

doi:10.18420/inf2019_31



(a) The Whiteside-data set and the main direction of interest.



(b) After the DipTransformation.

Fig. 2: We find the direction with the most structure (black: how much structure is found, grey: how large it is originally scaled) and scale the features according to it. It is now very easy to cluster.

distributed clusters. It should be very easy to cluster, but many algorithms (e.g. k-means) have massive difficulties with it, due to the scaling. Measuring the amount of structure of the features with the dip test and re-scaling the features leads to Fig. 1.b, a very easy to cluster data set. The heuristic here is that a feature is scaled relative to how much structure is found. The horizontal axis has barely any structure, it is uni-modal, and thus is scaled such that it has only a very small extension, which means that it has no great influence on clustering, as the values are all very similar. The feature with high structure – it is clearly multi-modal, i.e. has clusters one can distinguish from each other – is scaled such that this feature has a high influence on clustering.

The DipTransformation is not limited to axes-parallel re-scaling. Using a cleverly devised search strategy, it can automatically find non-axis-parallel features with high dip values, which it rescales as explained. One such example is shown in Fig. 2. The Whiteside-data set is a real-world data set that is difficult to handle for many clustering approaches, due to its clusters which are tricky to differentiate. Most approaches fail completely, but after the transformation, it is almost trivial.

In conclusion, we developed a technique that can improve the structure of a data set and thus its clustering. We show in [2] that this is true by testing it extensively on various data sets, all of which become far easier to cluster for various standard and state-of-the-art clustering approaches. DipTransformation assumes no data distribution, is deterministic, basically parameter-free and quite fast compared to various clustering approaches. It can thus be used as a pre-clustering step, that enhances the data set, and the clustering algorithm can be selected according to user preferences.

Bibliography

[1] Hartigan, J. A., Hartigan, P. M., *The Dip Test of Unimodality*, The Annals of Statistics, 1985.

[2] Schelling, B., Plant, C., *DipTransformation: Enhancing the Structure of a Dataset and Thereby Improving Clustering*, ICDM, 2018.

Discovering Non-Redundant K-means Clusterings in Optimal Subspaces (Extended Abstract)

Presentation of work originally published in the Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining

Dominik Mautz[1] Wei Ye[2] Claudia Plant[3] Christian Böhm[4]

Keywords: clustering; k-means; subspace; non-redundant

A huge object collection in high-dimensional space can often be meaningfully clustered in more than one way. For instance, objects could be clustered by their shape or alternatively by their color. Each grouping represents a different view of the data set. The new research field of *non-redundant clustering* addresses this type of problems. In our paper [Ma18], we follow the approach that different, non-redundant *k*-means-like clusterings may exist in different, arbitrarily oriented subspaces of the high-dimensional space. Minor assumptions about the orthogonality of the subspaces enable a particularly rigorous mathematical treatment of the non-redundant clustering problem and thus a particularly efficient algorithm, which we call Nr-Kmeans (for non-redundant *k*-means).

Figure 1 shows an example of a non-redundant clustering task. Given pictures of four objects—from the *Amsterdam Library of Object Images* (ALOI)— taken from different viewing angles and illumination temperatures could either be clustered by their shape into round and cylindrical objects or alternatively by their color into red and green objects. Each grouping represents a different view of the data set and is equally valid. From a mathematical perspective, there are two different low-dimensional subspaces, each exhibiting an interesting clustering structure. The clusterings in the subspaces are mutually non-redundant, i. e. each object belongs to different clusters in different subspaces. Classical clustering algorithms are not suited to capture these distinct views and may find only one of the possible partitions or a hard to interpret mixture of different clusterings.

The proposed non-redundant clustering algorithm Nr-Kmeans tackles this problem with a simple idea: find multiple mutually orthogonal subspaces—that may be arbitrarily oriented within the full space—such that the objective function of classical *k*-means is optimized in all of them. Both, the subspaces and the clusterings within are optimized simultaneously and influence each other during optimization. The only parameters needed are the expected number of clusters within each subspace. The orthogonality between subspaces ensures that the discovered clusterings represent different views on the data providing mutually

[1] Ludwig-Maximilians-Universität München, Munich, Germany, mautz@dbs.ifi.lmu.de
[2] University of California, Santa Barbara, CA, USA, weiye@cs.ucsb.edu
[3] Faculty of Computer Science, University of Vienna, Vienna, Austria
ds:UniVie, University of Vienna, Vienna, Austria, claudia.plant@univie.ac.at
[4] MCML, Ludwig-Maximilians-Universität München, Munich, Germany, boehm@dbs.ifi.lmu.de

doi:10.18420/inf2019_32

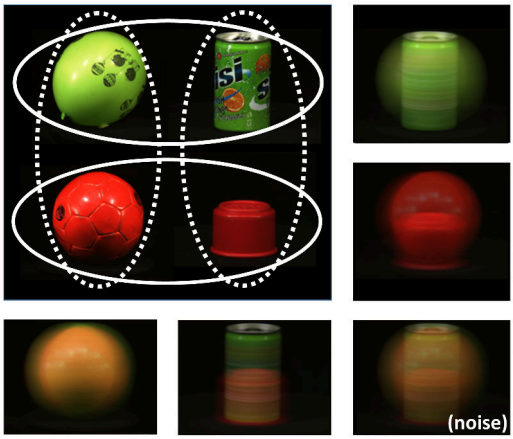


Fig. 1: Multiple clustering possibilities of objects according to color and shape. The smaller images show the corresponding average images.

non-redundant information and further allows for an efficient optimization procedure. The dimensionality of each subspace is determined automatically and the subspaces are well suited for visualization and further analysis, as they reveal the relationships between the individual clusters of a clustering. Inheriting from k-means, the result of Nr-Kmeans includes interpretable cluster centers, as displayed in Figure 1. In addition, our technique introduces a noise subspace, orthogonal to the other subspaces. The noise subspace captures all the unimodal variance in the data that is not interesting for clustering. This property allows Nr-Kmeans to prune away subspace dimensions without any clustering information and helps to outperform existing methods, especially on high-dimensional data.

Furthermore, it is possible to extend Nr-Kmeans with many other proposed k-means extensions in a straightforward manner, for instance, extensions exploiting the triangle inequality to speed up the assignment step, or we can simply initialize cluster centers within the subspaces using k-means++ or account for outliers with k-means--.

In our experiments, we show that Nr-Kmeans is a fast algorithm that, at the same time, yields results of a very high clustering quality with a high non-redundancy. Further, we show that the simultaneous optimization of both clustering and subspace is superior to an incremental clustering extraction procedure harnessed by some of the comparison methods. In short, we can say that Nr-Kmeans outperforms the comparison methods and discovers highly relevant combinations of subspaces and clusterings.

Bibliography

[Ma18] Mautz, Dominik; Ye, Wei; Plant, Claudia; Böhm, Christian: Discovering Non-Redundant K-means Clusterings in Optimal Subspaces. In (Guo, Yike; Farooq, Faisal, eds): Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2018, London, UK, August 19-23, 2018. ACM, pp. 1973–1982, 2018.

git2net: Mining Time-Stamped Co-Editing Networks from Large git Repositories

Presentation of work originally published in the Proc. of the 16th Intl. Conf. on Mining Software Repositories [GSS19]

Christoph Gote[1] Ingo Scholtes[2] Frank Schweitzer[3]

Keywords: repository mining; empirical software engineering; network science; data science; collaboration network; co-editing; social network analysis

Extended Abstract

Many software projects use version control systems like *git* to track changes in the developed source code. The analysis of collaboration networks constructed from co-editing relations stored in such *git* repositories can yield deep insights both into team-based software development processes as well as sociological theory. However, tools that allow to conveniently extract such rich, time-stamped collaboration networks for the large corpus of git repositories available are currently missing. Addressing this gap, the contributions of our work are as follows:

- We introduce `git2net`, an Open Source python tool that can be used to mine time-stamped and weighted co-editing relations between developers from the sequence of file modifications contained in *git* repositories. Here, each co-editing relationship $\big(A, B; t, w\big)$ represents developer A modifying w characters of code originally written by another developer B at time t. Utilising a parallel processing model the tool scales to massive software repositories with hundreds of thousands of commits and millions of lines of code.
- Analysing all file modifications contained in the *commit log*, `git2net` generates a database that captures fine-grained information on co-edited code either at the level of lines or contiguous code regions. It further analyses the overlap between co-edited code regions facilitating a character-based proxy estimating the effort behind code modifications. The approach is programming language agnostic.

[1] Chair of Systems Design, ETH Zürich, cgote@ethz.ch
[2] Data Analytics Group, Department of Informatics, University of Zürich, scholtes@ifi.uzh.ch
[3] Chair of Systems Design, ETH Zürich, fschweitzer@ethz.ch

doi:10.18420/inf2019_33

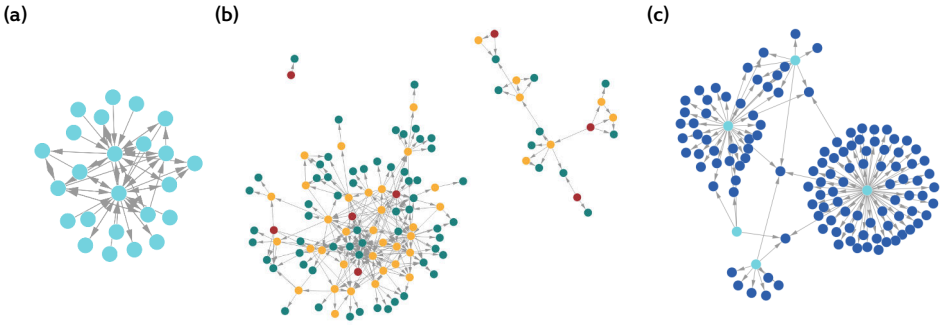


Fig. 1: Three time-aggregated collaboration networks generated by `git2net`. **(a)** Time-aggregated, static, directed network of co-editing relations. **(b)** Directed acyclic graph of edits of the a source code file. **(c)** Bipartite network linking developers (lightblue) to the files that they edited (blue).

- We develop methods to generate time-stamped collaboration networks based on multiple projections. Exemplary projections are shown in Figure 1.
- Applying `git2net` in a case study on two software projects, we show that the character-based analysis of file modifications yields considerably different network structures compared to previously used methods that have analysed code co-authorship at the level of files or modules. We further motivate how our tool can be used to address a set of research questions.

`git2net` facilitates the extraction of large-scale time-stamped network data that can be cross-referenced with project related information (e.g. project success, or organisational structures). We therefore expect the tool to be of considerable value for the network science community and researchers at the intersection of data science, computational science, and empirical software engineering.

Further details are available in the full paper [GSS19]. `git2net` is available as Open Source project on GitHub[4] and can be installed via `pip install git2net`. A tutorial reproducing the results from the full paper is available on `zenodo.org`[5].

References

[GSS19] Gote, Christoph; Scholtes, Ingo; Schweitzer, Frank: git2net: Mining Time-Stamped Co-Editing Networks from Large git Repositories. In: Proceedings of the 16th International Conference on Mining Software Repositories. MSR '19. IEEE Press, 2019.

[4] `https://github.com/gotec/git2net`
[5] 10.5281/zenodo.2587483

An Efficient Method for Exploratory Data Visualization of Big Spatial Data on Commodity Hardware

Presentation of work originally published in Geoinformatica (2019)

Christian Beilschmidt[1], Michael Mattig[1], Thomas Fober[1], Bernhard Seeger[1]

Abstract: The exploratory and interactive visualization of big spatial data is becoming increasingly important in business, science, and many other application areas. In this paper, we discuss the Circle Merging Quadtree, an efficient method for aggregating and visualizing big spatial point data on commodity hardware.

Keywords: Data Visualization, Biodiversity Data Analytics, Big Spatial Data Analysis

1 Summary

The exploratory and interactive visualization of big spatial data is becoming increasingly important in business, science, and many other application areas. For example, the demanding challenges in biodiversity requires new data-driven approaches to extract the information from an increasing amount of available heterogeneous data sources. Because scientific tasks generally result in complex analytical workflows with the necessity of having a researcher in the loop, efficient methods for the visualization of intermediate results are of utmost importance. While many of these methods are designed for dedicated hardware, there is currently a tendency that scientists make use of mobile devices instead. Thus, commodity hardware like tablets and smartphones are becoming the target end-user device for both, user interaction and scientific visualization. Therefore, methods for data visualization should not only offer low latency results, but also adapt to the limitations of the underlying devices like network bandwidth, battery power and screen resolution.

In this paper, we discuss an efficient method for aggregating and visualizing big spatial point data on commodity hardware. Our Circle Merging Quadtree (CMQ) method [Be19] offers a low latency, and thus supports users to explore big spatial data in an interactive manner. For two-dimensional point collections, our method uses a small set of non-overlapping circles such that (i) they follow the distribution of the points, (ii) they represent the cardinality of the underlying point subset by the circle area, (iii) they reveal hot spots while simultaneously keeping outliers. Fig. 1 exemplifies the transformation from raw points to aggregated circles.

[1] University of Marburg, Faculty of Mathematics and Computer Science, Hans-Meerwein-Straße, Marburg, Germany, {beilschmidt,mattig,thomas,seeger}@mathematik.uni-marburg.de

doi:10.18420/inf2019_34

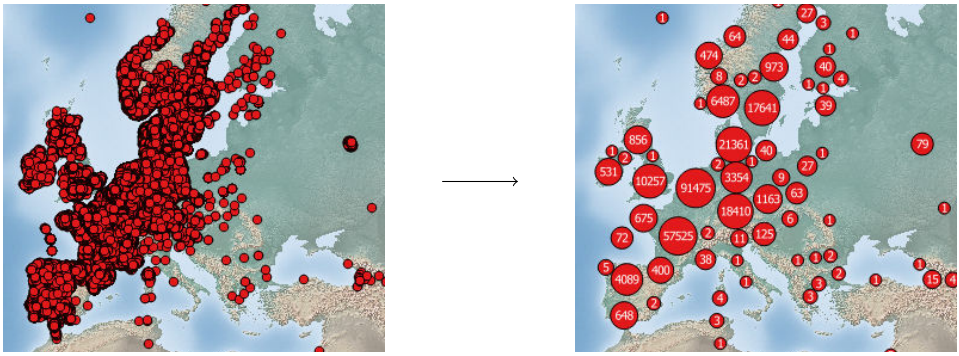


Fig. 1: This figure shows the discrepancy between mapping raw points (left) of the black alder and showing an aggregated view (right) of the points [Be17].

Based on a grid-based pre-aggregation step and the management of the resulting circles in a modified quadtree, our algorithm computes the final non-overlapping circles in linear time with respect to the number of points. Experimental results confirm its excellent runtime and quality in comparison to competitors. For instance, Jänicke et al. [JHS13], who provided one of the first algorithm to this problem that avoids overlaps, performs significantly slower in runtime complexity and practical performance.

2 Outlook

In geographic information systems, it is common to display multiple datasets as multiple layers on a map. This again leads to a cluttered representation with overlapping points and circles. As a future task, we will extend CMQ to process multiples point datasets at once into a joint representation. Furthermore, we aim at extending our method to compute aggregates of non-spatial attributes in the data.

References

[Be17] Beilschmidt, C.; Fober, T.; Mattig, M.; Seeger, B.: A Linear-Time Algorithm for the Aggregation and Visualization of Big Spatial Point Data. In: SIGSPATIAL '17. ACM, New York, NY, USA, 73:1–73:4, 2017.

[Be19] Beilschmidt, C.; Fober, T.; Mattig, M.; Seeger, B.: An Efficient Aggregation and Overlap Removal Algorithm for Circle Maps, GeoInformatica, 2019.

[JHS13] Jänicke, S.; Heine, C.; Scheuermann, G.: GeoTemCo: Comparative Visualization of Geospatial-Temporal Data with Clutter Removal Based on Dynamic Delaunay Triangulations. In: VISIGRAPP 2012. Vol. 359, Springer, Berlin, Heidelberg, Germany, pp. 160–175, 2013.

Serverless Big Data Processing using Matrix Multiplication as Example

Presentation of work originally published in IEEE Big Data 2018[We18]

Sebastian Werner[1] Jörn Kuhlenkamp[1] Markus Klems[2] Johannes Müller[3] Stefan Tai[1]

Keywords: serverless; big data; cloud services; matrix multiplication

Over the last years, performance and scalability needs for big data processing have been rather successfully addressed. This has been achieved by infrastructure platforms based on open-source distributed computing frameworks, e.g., Spark, TensorFlow, and Flink, that run on servers provisioned by cloud infrastructure services, e.g., AWS EC2. However, operating staff costs and infrastructure costs present significant cost factors for processing data-at-scale. Furthermore, processing big data on cloud infrastructure requires extensive knowledge to define and execute jobs and to deploy, configure, and maintain the required infrastructure platform for running jobs. Hence, the reduction of both costs and entry barriers for processing data-at-scale are grand challenges of data management [Ma18].

Serverless computing [Jo19] is emerging as a popular alternative model to on-demand cloud computing [Le18]. A client no longer uses cloud infrastructure directly, and instead only provides application logic that the serverless provider executes. While first prototypical implementations for serverless big data processing have been proposed [He16], developers and decision makers require hard evidence to make knowledgeable decisions about using traditional cloud infrastructure versus using FaaS. In our original research paper presented at the **IEEE BigData 2018 Conference** [We18], we address the research question of whether serverless platforms are feasible and beneficial for analyzing data-at-scale.

We build on our extensive expertise and previous work on quality-driven design and evaluation of cloud-based systems and answer the research question by means of experimental evaluation [BWT17; KW18]. We present generic requirements for designing serverless data processing applications, a prototype for distributed matrix multiplication, and extensive experiment results (see figure 1).

We showed that the utilization of serverless infrastructure indeed can lower operational and infrastructure costs without compromising established system qualities. Our experimental results indicate that serverless implementations can situationally compete, match, and

[1] Technische Universität Berlin, Information Systems Engineering, Germany, {sw, jk, st} @ise.tu-berlin.de
[2] WeAdvise AG, Munich Germany
[3] EXXETA AG, Berlin, Germany

doi:10.18420/inf2019_35

even outperform cluster-based distributed compute frameworks regarding performance and scalability. Furthermore, our approach enables developers to simply configure the cost/performance trade-off according to requirements at hand.Thus, a serverless solution can significantly reduce the entry barrier for new developers both in terms of costs and lowered complexity of configuring a sophisticated big data solution stack.

Fig. 1: Overview of our serverless processing architecture.

References

[BWT17] Bermbach, D.; Wittern, E.; Tai, S.: Cloud Service Benchmarking: Measuring Quality of Cloud Services from a Client Perspective. Springer, Cham, 2017.

[He16] Hendrickson, S.; Sturdevant, S.; Harter, T.; Venkataramani, V.; Arpaci-Dusseau, A. C.; Arpaci-Dusseau, R. H.: Serverless Computation with OpenLambda. In: 8th USENIX Workshop on Hot Topics in Cloud Computing. HotCloud'16, USENIX Association, Denver, CO, pp. 14–19, 2016.

[Jo19] Jonas, E.; Schleier-Smith, J.; Sreekanti, V.; Tsai, C.-C.; Khandelwal, A.; Pu, Q.; Shankar, V.; Carreira, J.; Krauth, K.; Yadwadkar, N.; Gonzales, J. E.; Popa, R. A.; Stoica, I.; Patterson, D. A.: Cloud Programming Simplified: A Berkeley View on Serverless Computing. CoRR/, pp. 1–33, Feb 2019.

[KW18] Kuhlenkamp, J.; Werner, S.: Benchmarking FaaS Platforms: Call for Community Participation. In: Proceedings of the 3rd International Workshop on Serverless Computing. WoSC'18, pp. 189–194, Dec. 2018.

[Le18] Leitner, P.; Wittern, E.; Spillner, J.; Hummer, W.: A Mixed-Method Empirical Study of Function-As-A-Service Software Development in Industrial Practice. Journal of Systems and Software 149/, pp. 340–359, Dec. 2018.

[Ma18] Markl, V.: Mosaics in Big Data: Stratosphere, Apache Flink, and Beyond. In. ACM Press, pp. 7–13, 2018, ISBN: 978-1-4503-5782-1.

[We18] Werner, S.; Kuhlenkamp, J.; Klems, M.; Müller, J.; Tai, S.: Serverless Big Data Processing using Matrix Multiplication as Example. In: 2018 IEEE International Conference on Big Data (Big Data). Pp. 358–365, Dec. 2018.

Mira: Sharing Resources for Distributed Analytics at Small Timescales

Presentation of work originally published in the Proc. of 2018 IEEE International Conference on Big Data, Seattle, USA

Michael Kaufmann[1,2] Kornilios Kourtis[1] Adrian Schuepbach[1] Martina Zitterbart[2]

Abstract: Mira is a system for optimized elastic execution of short-running and interactive data-analytics applications with low-latency execution startup, fast resource management and efficient resource utilization on shared clusters. We highlight the key insights and the Mira approach and summarize the most important results.

Keywords: Data Science; Distributed Analytics; Elastic Computing

1 Motivation

Modern distributed analytics stacks consist of application frameworks that enable processing of large amounts of data, and a resource manager that allows applications to share computational resources. These systems were designed to run batch jobs with long lifetimes (e.g., a few hours). New use cases, such as interactive applications, have emerged, which requires operating at smaller timescales (seconds) to share resources efficiently. Small timescales

(a) Spark+YARN (b) Spark+Mira (warm: executors are reused)

Fig. 1: Execution of a Spark application that spawns 110 tasks, each of which sleep for 10s.

pose a significant challenge for existing systems. To illustrate this, we run a simple Spark application on a YARN-managed cluster. The application spawns 110 tasks, each of which sleeps for 10 seconds. Fig. 1a shows the application's demands (blue line) and the resources that were allocated to it (red line). There is a significant delay in in the application acquiring the needed resources as well as in releasing them after it is done. For interactive applications, those costs dominate – or even exceed – the application runtime. Mira [Ka18] addresses

[1] IBM Research Zurich, Rueschlikon, Switzerland {kau,kou,dri}@zurich.ibm.com
[2] Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany zitterbart@kit.edu

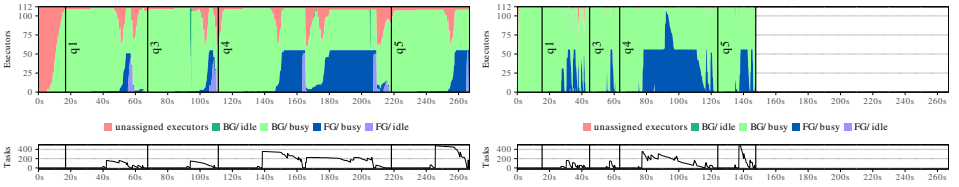
doi:10.18420/inf2019_36

this issue and enables resource management over small timescales (sub-seconds). Fig. 1b shows that the same application on Mira+Spark executes in significantly less time.

2 Mira

Mira includes two parts: a resource manager (RM) and an application scheduler (AS). The former manages all resources and applications, while the later schedules the tasks of a single application and communicates with the RM. Mira integrates with an existing application framework (AF) (for example Spark) and the AF's execution environment (EX). Mira achieves efficiency in two ways. First, it treats EX not as ephemeral, but as long-lived, shared resources. This allows it to minimize recurring acquisition costs and benefit from warmed-up executors (e.g., JIT, caches). Second, as consequence of the minimized resource acquisition cost, Mira is able to acquire and release resources almost instantaneously.

To evaluate Mira, we execute two applications concurrently. A background (BG) application, with an infinite loop of stages with 8192 1s-tasks per stage, generates a constant load on the cluster to force the RM to actively balance resource among applications. A foreground (FG) application runs TPC-DS queries. We compare to YARN as baseline. Overall, Mira reduces application runtime by up to 4.2× and ≈2.4× on average. Fig. 2 compares the resource allocation to the BG (green) and FG (blue) applications executed on YARN vs. Mira. Red areas represent executors not assigned to any application by the RM. Due to Mira's low executor assignment latency and shorter task runtime (size of blue areas), the FG on Mira is able to execute the same number of queries in 148s instead of 267s. Mira has virtually no unassigned executors (red spikes in Fig. 2b) during the resource-reassignment period. Finally, Spark+YARN suffers from the high cost of resource reacquisition. For query 4, Spark releases executors due to a short dip in task load (at ≈165s) just to reacquire them a few seconds later with a multi-second delay.



(a) Spark+YARN (b) Spark+Mira

Fig. 2: Resource sharing between BG and FG applications (top) and task load of the FG (bottom). FG application submissions are indicated by labelled vertical black lines.

Bibliography

[Ka18] Kaufmann, M. et.al.: Mira: Sharing Resources for Distributed Analytics at Small Timescales. In: Proceedings of the 2018 IEEE International Conference on Big Data. ACM, 2018.

On Coresets for Logistic Regression

Presentation of work originally published in Advances in Neural Information Processing Systems 31, NeurIPS 2018, [Mu18]

Alexander Munteanu[1] Chris Schwiegelshohn[2] Christian Sohler[3] David P. Woodruff[4]

Abstract: Coresets are one of the central methods to facilitate the analysis of large data. We continue a recent line of research applying the theory of coresets to logistic regression. First, we show the negative result that no strongly sublinear sized coresets exist for logistic regression. To deal with intractable worst-case instances we introduce a complexity measure $\mu(X)$, which quantifies the hardness of compressing a data set for logistic regression. $\mu(X)$ has an intuitive statistical interpretation that may be of independent interest. For data sets with bounded $\mu(X)$-complexity, we show that a novel sensitivity sampling scheme produces the first provably sublinear $(1 \pm \varepsilon)$-coreset.

Keywords: logistic regression, coresets, lower bounds, beyond worst-case analysis

1 Introduction

Scalability is one of the central challenges of modern data analysis and machine learning. Algorithms with polynomial running time might be regarded as efficient in a conventional sense, but nevertheless become intractable when facing massive data sets. As a result, performing data reduction techniques in a preprocessing step to speed up a subsequent optimization problem has received considerable attention. A natural approach is to sub-sample the data according to a certain probability distribution. In this paper we focus on the logistic regression problem which is an instance of a generalized linear model. We are given data $Z \in \mathbb{R}^{n \times d}$, and labels $Y \in \{-1, 1\}^n$. The optimization task consists of minimizing the negative log-likelihood $\sum_{i=1}^{n} \ln(1 + \exp(-Y_i Z_i \beta))$ with respect to the parameter $\beta \in \mathbb{R}^d$. To tackle scalability issues for logistic regression via sub-sampling we choose a probability distribution based on the *sensitivity* score of each point. Informally, the sensitivity of a point corresponds to the worst-case contribution of the point to the objective function we wish to minimize. If the total sensitivity, i.e., the sum of all sensitivity scores, is bounded by a reasonably small value, there exists a small collection of input points known as a *coreset* with very strong aggregation properties. For any solution $\beta \in \mathbb{R}^d$, the objective function evaluates on the coreset as on the original data up to a small multiplicative error [MS18].

[1] TU Dortmund, Data Science Center, 44221 Dortmund, Germany, alexander.munteanu@tu-dortmund.de
[2] Sapienza University of Rome, CS Department, 00185 Rome, Italy, schwiegelshohn@diag.uniroma1.it
[3] TU Dortmund, CS Department, 44221 Dortmund, Germany, christian.sohler@tu-dortmund.de
[4] Carnegie Mellon University, CS Department, Pittsburgh, PA 15213, USA, dwoodruf@cs.cmu.edu

doi:10.18420/inf2019_37

2 Our contributions

We show that logistic regression has no sublinear streaming algorithm. Due to a standard reduction between coresets and streaming algorithms, this implies that logistic regression admits no sublinear coresets or bounded sensitivity scores in general.

We investigate available sensitivity sampling distributions for logistic regression. For points with large contribution, where $-Y_i Z_i \beta \gg 0$, the objective function increases by a term almost linear in $-Y_i Z_i \beta$. This motivates to use sensitivity scores designed for ℓ_1-related problems. To this end, we propose sampling from a mixture distribution with one component proportional to the *square root* of the ℓ_2^2 leverage scores. The other mixture component is uniform sampling to deal with the remaining domain. Our experiments show that this distribution outperforms uniform and k-means based sensitivity sampling by a wide margin on real data sets. The algorithm is space efficient, and can be implemented in a variety of models used to handle large data sets such as 2-pass streaming, and massively parallel frameworks such as Hadoop and MapReduce, and can be implemented to work in input sparsity time, i.e., proportional to the number of non-zero entries of the data [Wo14].

We analyze our sampling distribution for a parametrized class of instances we call μ-complex, placing our work in the framework of *beyond worst-case analysis* [Ro19]. The parameter μ roughly corresponds to the ratio between the log of correctly estimated odds and the log of incorrectly estimated odds. The condition of small μ is justified by the fact that for instances with large μ, logistic regression exhibits methodological problems. We show that the total sensitivity of logistic regression can be bounded in terms of μ. Moreover, if the data is μ-complex for a small, not necessarily constant μ, then there exists a sampling and reweighting scheme based on the sensitivity framework that produces a $(1 \pm \varepsilon)$-coreset of sublinear size $O(\varepsilon^{-2}\mu\sqrt{n}d^{3/2}\log^2(\mu nd))$ with high probability. A more involved recursive sampling scheme produces a $(1 \pm \varepsilon)$-coreset of size $O(\varepsilon^{-4}\mu^3 d^3 \log^{O(1)}(\mu nd))$, which is beneficial if the data is well-behaved and the input size is particularly large. These are the first provably sublinear coreset constructions for logistic regression.

Bibliography

[MS18] Munteanu, Alexander; Schwiegelshohn, Chris: Coresets-Methods and History: A Theoreticians Design Pattern for Approximation and Streaming. KI, 32(1):37–53, 2018.

[Mu18] Munteanu, Alexander; Schwiegelshohn, Chris; Sohler, Christian; Woodruff, David P.: On Coresets for Logistic Regression. In: Advances in Neural Information Processing Systems (NeurIPS). pp. 6562–6571, 2018.

[Ro19] Roughgarden, Tim: Beyond worst-case analysis. Commun. ACM, 62(3):88–96, 2019.

[Wo14] Woodruff, David P.: Sketching as a Tool for Numerical Linear Algebra. Foundations and Trends in Theoretical Computer Science, 10(1-2):1–157, 2014.

Priors for Linear Differential Equations

Presentation of work originally published in Advances in Neural Information Processing Systems 31 (NeurIPS 2018)

Markus Lange-Hegermann[1]

Abstract: We algorithmically construct multi-output Gaussian process priors which satisfy linear differential equations. We parametrize all solutions of the differential equations using Gröbner bases for controllable systems. If successful, a push forward along the parametrization is the desired prior. This prior yields an interpretable machine learning model, which can combine linear differential equations with noisy data points.

Keywords: Gaussian process; regression; differential equation; kernel; Gröbner basis

In recent years, Gaussian process regression has emancipated itself from pure math research and has become a prime regression technique in machine learning and data science. Roughly, a Gaussian process can be viewed as a suitable probability distribution on a set of functions, which we can condition on observations using Bayes' rule. The resulting mean function is used for regression. The strength of Gaussian process regression lies in *avoiding overfitting* while still finding functions complex enough to describe *any behavior* present in given observations, even in noisy or unstructured data. Gaussian processes are usually applied when data is rare or expensive to produce.

Incorporating justified assumptions into the prior helps in applications: the full information content of the scarce observations can be utilized to create a more precise regression model. Furthermore, such assumptions allow for an interpretable model. Examples of such assumptions are smooth or rough behavior, trends, and periodicity. Such assumptions are usually incorporated in the covariance structure of the Gaussian process.

As Gaussian processes are, roughly speaking, the linear objects among stochastic processes, specific linear differential equations could be incorporated into the covariance structures of Gaussian processes, see e.g. [MC08, So15]. A first step towards systematizing this construction was achieved in [Ji17], where in special cases, a map φ into the solution set for physical laws could be found. With φ, one could assume a Gaussian process prior in its domain and push it forward. This results in a Gaussian process prior for the solutions of the physical laws. However, the approach of [Ji17] to compute φ does not necessarily terminate.

[1] OWL University of Applied Sciences and Arts, Department of Electrical Engineering and Computer Science, Campusallee 12, 32657 Lemgo, Germany, markus.lange-hegermann@th-owl.de

doi:10.18420/inf2019_38

This is an extended abstract of [LH18]. In that paper, we stress that the map φ from [Ji17] needs to be a parametrization. This allows to use Gröbner bases algorithms to compute φ if it exists or report failure if it does not exist. It turns out that the approach works for controllable system, but no similar Gaussian priors can exist for non-controllable systems. We furthermore extend the approach from linear differential equations with constant coefficients to variable coefficients.

Using this model, one can add information to Gaussian processes[2] not only by

(i) conditioning on observations (Bayes' rule), but also by

(ii) restricting to solutions of linear operator matrices by constructing a suitable prior.

Since these two constructions are compatible, we can combine *strict, global information* from equations with *noisy, local information* from observations.

This combination of techniques yields a grey box model without any numerical solving or approximation of differential equations. In fact, we achieve a non-parametric regression that can yield only solutions of the system of differential equations, but also all solutions. As customary with Gaussian processes, one can of course take noise into consideration.

In [LH18] we show how to translate controlling a dynamical linear systems into a regression problem and apply data-driven methods to control such systems. Furthermore, we construct a Gaussian process such that all of its realizations satisfy the inhomogeneous Maxwell equations of electromagnetism. We show how to extrapolate away from the data by conditioning this Gaussian process on a *single* observation of electric current, which yields as expected a magnetic field circling around this electric current. Finally, we construct a Gaussian process prior for vector fields on the surface of the sphere. These vector fields can additionally be constrained to be divergence free fields, i.e., without sources and sinks.

Bibliography

[Ji17] Jidling, Carl; Wahlström, Niklas; Wills, Adrian; Schön, Thomas B.: Linearly Constrained Gaussian Processes. 2017. (arXiv:1703.00787).

[LH18] Lange-Hegermann, Markus: Algorithmic Linearly Constrained Gaussian Processes. In Advances in Neural Information Processing Systems 31, pp. 2137–2148. Curran Associates, Inc., 2018.

[MC08] Macêdo, Ives; Castro, Rener: Learning divergence-free and curl-free vector fields with matrix-valued kernels. Instituto Nacional de Matematica Pura e Aplicada, Brasil, 2008.

[So15] Solin, Arno; Kok, Manon; Wahlström, Niklas; Schön, Thomas B.; Särkkä, Simo: Modeling and interpolation of the ambient magnetic field by Gaussian processes. 2015. (arXiv:1509.04634).

[2] The construction of covariance functions is applicable to kernels more generally.

On Advancement of Information Spaces to Improve Prediction-Based Compression

Presentation of work originally published in the Proc. of the 2018 IEEE International Conference on Big Data

Ugur Cayoglu[1] Frank Tristram[2] Jörg Meyer[1] Tobias Kerzenmacher[3] Peter Braesicke[3] Achim Streit[1]

Abstract: One of the scientific communities that generate the largest amounts of data today are the climate sciences. New climate models enable model integrations at unprecedented resolution, simulating timescales from decades to centuries of climate change. Nowadays, limited storage space and ever increasing model output is a big challenge. For this reason, we look at lossless compression using prediction-based data compression. We show that there is a significant dependence of the compression rate on the chosen traversal method and the underlying data model. We examine the influence of this structural dependency on prediction-based compression algorithms and explore possibilities to improve compression rates. We introduce the concept of Information Spaces (IS), which help to improve the accuracy of predictions by nearly 10% and decrease the standard deviation of the compression results by 20% on average.

Keywords: compression algorithms; encoding; meteorology; prediction-based compression; information spaces

Introduction

New climate models such as ICON-ART [Sc18] make it possible to run high-resolution simulations of the atmosphere and its composition at an unprecedented scale and detail while making full use of the available capacity of high-performance computers. But with these improvements, the storage space required to save the output of the simulations also increases. In such situations an efficient compression method can help reduce the required storage space. In prediction-based compression all data points are processed in a predefined traversal sequence. As the sequence is processed, a prediction is given for each data point based on the values prior in the sequence. The difference between the actual value and its prediction will then be saved on disk. A good prediction leads to a better compression rate. For an extended version of this work please refer to our original publication [Ca18].

[1] Karlsruhe Institute of Technology, Steinbuch Centre for Computing (SCC), Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany, Ugur.Cayoglu@kit.edu
[2] Karlsruhe Institute of Technology, Institute of Applied Physics (APH)
[3] Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-ASF)

doi:10.18420/inf2019_39

Method

Our method calculates position and neighbourhood information of each data point s_i during its traversal. We call this the Information Space (IS) of the data point.

Let S_i be all the data points in the data ordered by traversal path. The IS of a data point s_i is the set of data points $s_j \in S_i$ with $j < i$ and each element of the coordinate tuple within a certain range r of s_i.

$$IS(s_i) = \{s_k | \forall s_k \in S_i : a_j^i - r \leq a_j^k \leq a_j^i + r\} \tag{1}$$

with a_m^n defining the coordinate position at dimension m of element n of sequence S. The IS is then divided into its components to isolate the information contained in the various dimensions. We call these components Information Context (IC).

Each IC contains information along several dimensions. ICs can contain overlapping data points, but none is a subset of another. This allows predictions on the basis of information from different dimensions and later merge them into a consolidated prediction.

Evaluation

We analysed the performance of different prediction-based compression algorithms on climate data. The results of our experiments show that changing the starting point of the compression algorithm has only negligible effects on the compression rate, while changing the traversal path can influence the compression rate significantly. Further experiments show that with the help of IS it is possible to improve the predictions of each predictor. More importantly, the stability of the predictions can be increased. This results in higher quality forecasts with less fluctuations than with established methods.

Our current configuration achieved a 10% improvement in prediction accuracy and decreased the standard deviation of the compression results by over 20% on average.

Bibliography

[Ca18] Cayoglu, U.; Tristram, F.; Meyer, J.; Kerzenmacher, T.; Braesicke, P.; Streit, A.: Concept and Analysis of Information Spaces to improve Prediction-Based Compression. In: 2018 IEEE International Conference on Big Data (Big Data). pp. 3392–3401, Dec 2018.

[Sc18] Schröter, J.; Rieger, D.; Stassen, C.; Vogel, H.; Weimer, M.; Werchner, S.; Förstner, J.; Prill, F.; Reinert, D.; Zängl, G.; Giorgetta, M.; Ruhnke, R.; Vogel, B.; Braesicke, P.: ICON-ART 2.1 – A flexible tracer framework and its application for composition studies in numerical weather forecasting and climate simulations. Geoscientific Model Development Discussions, 2018:1–37, 2018.

From Automated to On-The-Fly Machine Learning

Presentation of work originally published in Machine Learning 107

Felix Mohr[1] Marcel Wever[1] Alexander Tornede[1] Eyke Hüllermeier[1]

Keywords: automated machine learning; classification; on-the-fly computing

Automated machine learning (AutoML) is the task of automatically selecting and parametrizing machine learning algorithms, as well as combining them into an overall solution (a "machine learning pipeline") specifically tailored for a task at hand (typically specified by a dataset). Existing approaches to AutoML are based on Bayesian optimization (e.g. auto-sklearn [Fe15]) or genetic algorithms (e.g. TPOT [Ol16]).

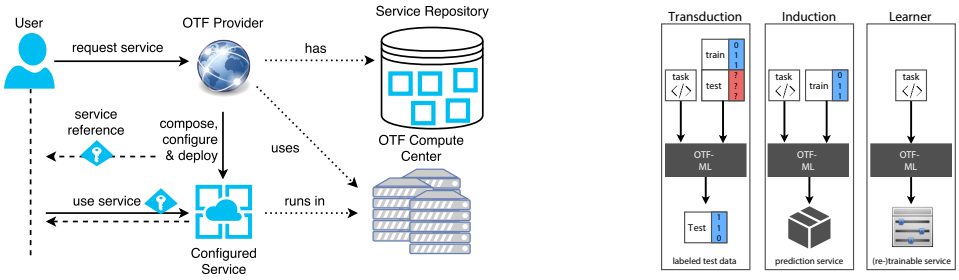
We recently complemented the repertoire of state-of-the-art AutoML tools by ML-Plan [MWH18b, WMH18, We19]. ML-Plan leverages techniques from hierarchical task network (HTN) planning to arrange the more than 10^{40} different candidate pipelines in a tree-shaped search space. In an extensive series of experiments, we showed that ML-Plan is highly competitive and often outperforms existing approaches.

Building on ML-Plan, our current work is devoted to the vision of what we call "On-the-Fly Machine Learning" (OTF-ML) — an instantiation of the On-the-Fly (OTF) computing paradigm [Ha13] for the case of machine learning, and, as such, an extension of AutoML. OTF computing aims at the provision of individually configured IT services in a dynamic, distributed market environment, which comprises different types of agents and allows customers to request services specifically tailored for their needs (cf. Fig. 1a).

In OTF-ML, we distinguish three types of services a customer may request (cf. Fig. 1b). In the *Transduction* scenario, the customer is interested in automatically labeling data. To this end, he provides a task description, along with training data and the data to be labeled. Internally, the OTF provider configures an ML service with the help of the provided training data and returns the labels for the unlabeled data obtained by the ML service. In the *Induction* scenario, the request only specifies the task and the training data. The customer is then provided access to the configured ML service, which can be queried to make predictions for new data points. Lastly, in the *Learner* scenario, the customer only describes the type of ML problem to be solved. He then obtains an ML service specifically tailored for such problems, which can be used for learning on whatsoever training data.

[1] Paderborn University, Warburger Straße 100, 33098 Paderborn, Germany <firstname>.<lastname>@upb.de

doi:10.18420/inf2019_40



a) Process of OTF provision of customized services

b) On-the-fly ML scenarios

Fig. 1: OTF-ML: the on-the-fly selection, configuration, provision, and execution of machine learning and data analytics functionality as requested by an end-user.

Realizing automated machine learning in an OTF environment offers various opportunities, including better computational resources, high parallelization and the combination of algorithms implemented for different platforms. In [Mo18, MWH18a], we extended ML-Plan to work on a service level combining implementations across plattforms. We consider these as important first steps paving the way for OTF-ML.

References

[Fe15] Feurer, M.; Klein, A.; Eggensperger, K.; Springenberg, J. T.; Blum, M.; Hutter, F.: Efficient and Robust Automated Machine Learning. In: Advances in Neural Information Processing Systems 28. 2015.

[Ha13] Happe, M.; Meyer auf der Heide, F.; Kling, P.; Platzner, M.; Plessl, C.: On-The-Fly Computing: A Novel Paradigm for Individualized IT Services. In: 16th Int. Symposium on Object/Component/Service-Oriented Real-Time Distributed Computing. 2013.

[Mo18] Mohr, F. Wever, M. Hüllermeier, E. Faez, A. Towards the Automated Composition of Machine Learning Services. In: IEEE International Conference on Services Computing. 2018.

[MWH18a] Mohr, F.; Wever, M.; Hüllermeier, E.: Automated Machine Learning Service Composition. CoRR, abs/1809.00486, 2018.

[MWH18b] Mohr, F.; Wever, M.; Hüllermeier, E.: ML-Plan: Automated Machine Learning via Hierarchical Planning. Machine Learning, 107(8-10):1495–1515, 2018.

[Ol16] Olson, R. S.; Urbanowicz, R. J.; Andrews, P. C.; Lavender, N. A.; Kidd, L. C.; Moore, J. H.: Automating Biomedical Data Science Through Tree-Based Pipeline Optimization. Applications of Evolutionary Computation, 2016.

[We19] Wever, M.; Mohr, F.; Tornede, A.; Hüllermeier, E.: Automating Multi-Label Classification Extending ML-Plan. 6th ICML Workshop on Automated Machine Learning, 2019.

[WMH18] Wever, M.; Mohr, F.; Hüllermeier, E.: ML-Plan for Unlimited-Length Pipelines. 5th ICML Workshop on Automated Machine Learning, 2018.

Tensor Methods for Global Sensitivity Analysis

Presentation of work originally published in the Journal of Reliability Engineering and System Safety, as well as in the SIAM/ASA Journal on Uncertainty Quantification

Rafael Ballester-Ripoll[1] Enrique G. Paredes[2] Renato Pajarola[3]

Abstract: Sobol indices and other, more recent quantities of interest (such as the effective and mean dimensions, the dimension distribution, or the Shapley values) are of great aid in sensitivity analysis, uncertainty quantification, and model interpretation. Unfortunately, computing such indices is still challenging for high-dimensional systems. We propose the tensor train decomposition (TT) as a unified framework for surrogate modeling and sensitivity analysis of independently distributed variables. To this end, we introduce the Sobol tensor train (Sobol TT) data structure, which compactly represents variance components for all possible joint variable interactions of any order. Our formulation allows efficient aggregation and subselection operations, and we are able to obtain related Sobol indices and other related quantities at low computational cost.

Keywords: sensitivity analysis; sobol indices; surrogate modeling; data visualization; multidimensional data analytics; tensor approximation

1 Motivation

Many models in computational science admit the general form $f(x_1, \ldots, x_N) \to \mathbb{R}$, given by an analytical formula or an approximator (a regressor) trained over some sample set. Variance-based sensitivity analysis (SA), and *Sobol's method* [Sa04] are powerful tools for model interpretation, where the *Sobol indices* measure the variance in f that is due to each individual input $x_1, \ldots, x_N$, as well as any combinations (interactions). Such interactions are crucial to gain insights on the underlying model, as some variables may be important only for certain values of one or more of the remaining variables. Among many other applications, Sobol indices can detect irrelevant variables or variables whose effect can be separated, e.g. $f(x_1, x_2, \ldots) = g(x_1) + h(x_2, \ldots)$. Those indices, however, arise from multidimensional integrals and are thus computationally challenging.

[1] Swiss Federal Institute of Technology Zurich, Switzerland barafael@ethz.ch
[2] Swiss National Supercomputing Center enrique.gonzalez@cscs.ch
[3] Department of Informatics, University of Zurich, Switzerland pajarola@ifi.uzh.ch

doi:10.18420/inf2019_41

2 Tensor-based Sensitivity Analysis

In order to conduct effective and affordable SA, we propose to use *tensor decompositions*, i.e. the *tensor train* (TT) model [Os11]. It is a numerical framework that scales well with the number of dimensions, and operates on the assumption that the model of interest f can be well-approximated by a *low-rank tensor* $\mathcal{T}$ (a *surrogate model*): $f(x_1, \ldots, x_N) \approx \mathcal{T}[i_1, \ldots, i_N]$. The low-rank assumption generalizes the notion of low-rank matrix factorization to three and more dimensions and entails two benefits: it acts as a regularization prior during model fitting, and it drastically reduces the computational cost of model postprocessing. Fortunately, a wide class of functions are well-approximable by low-rank tensor formats.

We first obtain an approximation $\mathcal{T}$ of a function f using the *cross-approximation* sampling method [OT10], then derive a new *Sobol tensor train* $\mathcal{S}$ which is the first data structure that gathers *all* 2^N Sobol indices of an N-dimensional model. In addition, the Sobol TT can be further manipulated in compact form to (a) extract more advanced sensitivity metrics and statistics, and (b) answer a wide range of queries including e.g. *What are the k most important variables?*, *What variable interacts the most with* x_n*?*, and many others. To satisfy such queries we propose a new class of so-called *tensor automata*, which are able to select and combine SA indices in many possible ways. For example, the *Hamming mask tensor* computes the combined importance of all k-plets of variables.

This extended abstract summarizes our work from [BRPP18] and [BRPP19].

Acknowledgments

This work was partially supported by the UZH Forschungskredit “Candoc”, grant number FK-16-012.

Bibliography

[BRPP18] Ballester-Ripoll, Rafael; Paredes, Enrique G.; Pajarola, Renato: Tensor Algorithms for Advanced Sensitivity Metrics. SIAM/ASA Journal on Uncertainty Quantification, 6(3):1172–1197, 2018.

[BRPP19] Ballester-Ripoll, Rafael; Paredes, Enrique G.; Pajarola, Renato: Sobol Tensor Trains for Global Sensitivity Analysis. Reliability Engineering and System Safety, 183:311–322, March 2019.

[Os11] Oseledets, Ivan V.: Tensor-Train Decomposition. SIAM Journal on Scientific Computing, 33(5):2295–2317, September 2011.

[OT10] Oseledets, I. V.; Tyrtyshnikov, E. E.: TT-cross Approximation for Multidimensional Arrays. Linear Algebra Applications, 432(1):70–88, 2010.

[Sa04] Saltelli, Andrea; Tarantola, Stefano; Campolongo, Francesca; Ratto, Marco: Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models. Halsted Press, New York, NY, USA, 2004.

A New Approach for Automated Feature Selection

Presentation of work originally published in the Proc. of the 2018 IEEE International Conference on Big Data [GLS18]

Andreas Gocht[1], Christoph Lehmann[1], Robert Schöne[1]

Keywords: Data mining; Feature selection; Mutual information

While more and more data is collected for machine learning, validation and exploration of these data become increasingly challenging. Moreover, if the collection of feature data is expensive, or the amount of usable features is limited, *feature selection* is often employed.

In [GLS18] we present an algorithm, which is able to select the most relevant features and stops automatically once the information added does not improve the quality of the result anymore. It is possible to specify an upper limit of features, and our algorithm is independent of any following machine learning task. The selection algorithm is based on the so-called Historical JMI (HJMI) score, as it uses the information from already selected features:

$$J_{HJMI}(X_k, S) = J_H + I(X_k; Y) - \frac{\sum_{X_j \in S}[I(X_k; X_j) - I(X_k; X_j|Y)]}{|S|}.$$

The set S contains already selected features. X_k refers to the currently investigated feature and X_j to an already selected feature out of S. X_k and X_j are features out of $X = \{X_1, X_2, ..., X_l\}$, where l denotes the number of all features. Y defines the target variable, which we like to predict. J_H is the historical information about the selected features, $I(X_k; Y)$ and $I(X_k; X_j)$ refers to the mutual information. $I(X_k; X_j|Y)$ specifies the conditional mutual information.

The algorithm to calculate the HJMI is given by:

1. Set $J_H = 0$
2. Calculate $J_{HJMI}(X_k, S)$ for all $X_k \in X \backslash S$
3. Save the largest result for $J_{HJMI}(X_k, S)$ as J_H and add the associated X_k to S
4. Repeat 2 and 3 until the stopping criterion is met or the maximum amount of features is reached

[1] Center for Information Services and High Performance Computing (ZIH) Technische Universität Dresden, 01062 Dresden, Germany, {andreas.gocht|christoph.lehmann|robert.schoene}@tu-dresden.de

doi:10.18420/inf2019_42

As for stopping criterion, we propose $\delta > \frac{J_{HMI}-J_H}{J_H}$, i.e., if the information added by a new feature X_k does not increase the information of the already selected features in S by more than a given δ, the algorithm will stop. However, as in a typical exploratory setting, the choice of the stopping criterion is to be reflected case-dependent as well as its interpretation. Compared to PCA, which only considers the pairwise dependence of features, the JMI is more general as it approximates the mutual information of the conditional distribution $Y|(S)$.

To evaluate our approach, we used the NIPS Feature Selection Challenge [Gu04], similar to [Br12]. The results are shown in Table 1. A detailed analysis can be found in [GLS18].

Benchmark	JMI Validation Error [%]	JMI Amount of Features (l)	HJMI Validation Error [%]	HJMI Amount of Features (l)
ARCENE	21.19	20	19.64	32
DEXTER	15.0	60	13.0	21
DOROTHEA	32.99	200	25.63	24
GISETTE	4.1	200	8.0	26
MADELON	10.67	20	10.67	20

Tab. 1: Results for NIPS Feature Selection Challenge. The first column shows the smallest error for the validation set, with features selected using JMI. The second column shows the amount of features used to achieve this result. The third and fourth column present the same information for the newly introduced HJMI-based algorithm. In most cases, HJMI is as good or better than the JMI. For GISETTE, JMI outperforms HJMI. However, it depends on the application if a reduction of only 3.9% in prediction error is worth selecting 174 additional features.

Acknowledgments This work is supported by the European Union's Horizon 2020 program in the READEX project (grant agreement number 671657) and the German Federal Ministry of Education and Research (BMBF, 01IS14014A-D) by funding the competence center for Big Data "ScaDS Dresden/Leipzig".

References

[Br12] Brown, G.; Pocock, A.; Zhao, M.-J.; Luján, M.: Conditional Likelihood Maximisation: A Unifying Framework for Information Theoretic Feature Selection. The Journal of Machine Learning Research/, ACMID: 2188387, 2012.

[GLS18] Gocht, A.; Lehmann, C.; Schöne, R.: A New Approach for Automated Feature Selection. In: 2018 IEEE International Conference on Big Data (Big Data). DOI: 10.1109/BigData.2018.8622548, 2018.

[Gu04] Guyon, I.; Gunn, S. R.; Ben-Hur, A.; Dror, G.: Result Analysis of the NIPS 2003 Feature Selection Challenge. In: Advances in Neural Information Processing Systems. ACMID: 2976109, 2004.

A Data Science Perspective on Deconvolution

Presentation of work originally published in the Proc. of the 5th Int. Conf. on Data Science and Advanced Analytics (DSAA)

Mirko Bunse[1], Nico Piatkowski[1], Tim Ruhe[2], Katharina Morik[1], Wolfgang Rhode[2]

Keywords: deconvolution; unfolding; supervised learning; transductive learning; density estimation; Cherenkov astronomy; regularization

Deconvolution problems arise when the probability density function (pdf) of a quantity Y is estimated even though Y cannot be measured directly. In this scenario, the pdf of Y has to be inferred from related quantities $X_1, X_2, \ldots$ which are measured instead. Several algorithms solving this task have been proposed in particle physics, a field where deconvolution problems frequently arise. In this extended abstract, we summarize our findings made from a data science perspective [Bu18] and our on-going work on deconvolution.

The term "de-convolution" (also known as "unfolding") is motivated by the traditional formalization of the problem, which models the pdf $g : \mathcal{X} \to \mathbb{R}$ of the observed quantities as a convolution of the sought-after pdf $f : \mathcal{Y} \to \mathbb{R}$ of Y with another function $R : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$. In this model, R represents a conditional probability function which is learned from a set of training data. The goal is to infer f from given g and R.

$$g(\vec{x}) \;=\; \int_{\mathcal{Y}} R\,(\vec{x} \mid y) \cdot f\,(y)\;\mathrm{d}y \tag{1}$$

Traditional approaches maximize the likelihood [Bl85] or employ Bayes' theorem [D'95] in a discrete variant of this formalization. Thus, they estimate the probability $\mathbb{P}(Y \equiv i)$ of each discrete state i of Y. Unfortunately, previous publications only present these methods as single monolithic instances. Two of our contributions are the unification of traditional algorithms and the identification of theoretic similarities between them.

Furthermore, we advocate a more recent approach [Ru16] based on supervised machine learning. It recasts deconvolution as a classification task, providing a modular framework in which the learning method is exchangeable. The idea is to recover each $\mathbb{P}(Y \equiv i)$ from a classifier's confidence $c_{\mathcal{M}}\,(i \mid \vec{x})$, which is interpreted as a probability conditioned on each observation. This reconstruction is then repeated in an expectation maximization (EM) procedure. While the original algorithm exhibits a diverging behavior, we propose several improvements which lead to a robust and also accelerated algorithm.

[1] TU Dortmund, Artificial Intelligence Group, D-44227 Dortmund, firstname.lastname@tu-dortmund.de
[2] TU Dortmund, Astroparticle Physics Group, D-44227 Dortmund, firstname.lastname@tu-dortmund.de
This work has been supported by Deutsche Forschungsgemeinschaft (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Data Analysis", projects C3 and A1. https://sfb876.tu-dortmund.de.

doi:10.18420/inf2019_43

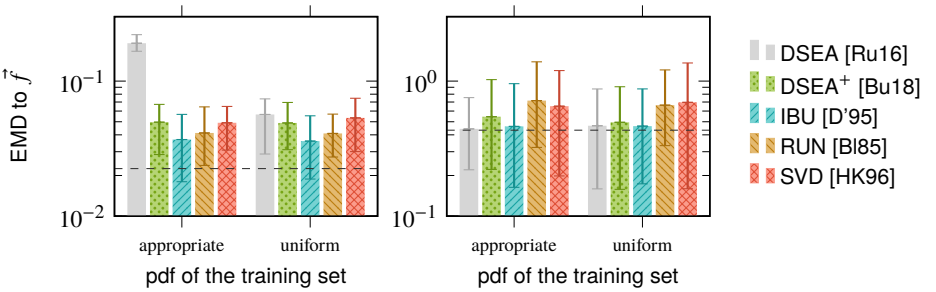


Fig. 1: The reconstruction quality of five methods is assessed in terms of the Earth Mover's Distance (EMD) between their estimates and the true discrete solution $\vec{f} \in \mathbb{R}^d$. On the left, we present an experiment with many observations [Bu18], while on the right only few observations are available for deconvolution (on-going work). Our improved method, DSEA$^+$, performs well throughout.

$$\hat{\mathbb{P}}(Y \equiv i) \;=\; \sum_{\vec{x} \in X} \hat{\mathbb{P}}(Y \equiv i | X = \vec{x}) \cdot \hat{\mathbb{P}}(X = \vec{x}) \;=\; \sum_{n} c_{\mathcal{M}}(i \mid \vec{x}_n) \cdot \frac{1}{N} \tag{2}$$

Finally, we evaluate the traditional approaches and the learning-based method in comparative experiments. The essence of our findings, as indicated by Fig. 1, is that all methods are able to obtain results of a similarly high quality, given that they are provided enough observations. One notable exception to this end is the learning-based approach *without* our improvements, which produces less accurate results.

In our on-going work, we are investigating relations between deconvolution and other tasks in machine learning. For example, we establish a connection to transductive learning. Also, we apply our methods to text corpora of political manifestos, thus demonstrating that deconvolution is a general data science problem by far not limited to particle physics.

References

[Bl85] Blobel, V.: Unfolding methods in high-energy physics experiments. Technical report, CERN, 1985.

[Bu18] Bunse, Mirko; Piatkowski, Nico; Ruhe, Tim; Rhode, Wolfgang; Morik, Katharina: Unification of Deconvolution Algorithms for Cherenkov Astronomy. In: Proc. of the 5th DSAA. IEEE, pp. 21–30, 2018.

[D'95] D'Agostini, G.: A multidimensional unfolding method based on Bayes' theorem. Nucl. Instrum. Methods Phys. Res. A, 362(2-3):487–498, 1995.

[HK96] Hoecker, A.; Kartvelishvili, V.: SVD approach to data unfolding. Nucl. Instrum. Methods Phys. Res. A, 372(3):469–481, 1996.

[Ru16] Ruhe, T.; Börner, M.; Wornowizki, M. et al.: Mining for Spectra – The Dortmund Spectrum Estimation Algorithm. In: Proc. of the ADASS XXVI. 2016. in press.

Safe Active Learning for Time-Series Modeling with Gaussian Processes

Presentation of work originally published in the Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada

Christoph Zimmer[1], Mona Meister[2], Duy Nguyen-Tuong[3]

Abstract: Learning time-series models is useful for many applications, such as simulation and forecasting. In this study, we consider the problem of actively learning time-series models while taking given safety constraints into account. For time-series modeling we employ a Gaussian process with a nonlinear exogenous input structure. The proposed approach generates data appropriate for time series model learning, i.e. input and output trajectories, by dynamically exploring the input space. The approach parametrizes the input trajectory as consecutive trajectory sections, which are determined stepwise given safety requirements and past observations. We analyze the proposed algorithm and evaluate it empirically on a technical application. The results show the effectiveness of our approach in a realistic technical use case.

Presentation of work [ZMNT18] originally published in the Proceedings of the 32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

Keywords: Safe Active Learning; Dynamics Modeling

Bibliography

[ZMNT18] Zimmer, Christoph; Meister, Mona; Nguyen-Tuong, Duy: Safe Active Learning for Time-Series Modeling with Gaussian Processes. In: 32nd Conference on Neural Information Processing Systems. 2018.

[1] Bosch Center for Artificial Intelligence, Renningen, Germany christoph.zimmer@de.bosch.com
[2] Bosch Center for Artificial Intelligence, Renningen, Germany mona.meister@de.bosch.com
[3] Bosch Center for Artificial Intelligence, Renningen, Germany duy.nguyen-tuong@de.bosch.com

doi:10.18420/inf2019_44

Knowledge-Based Short Text Categorization Using Entity and Category Embedding

Presentation of work originally published in the Proc. of the 16th ESWC 2019

Rima Türker[1] Lei Zhang[1] Maria Koutraki[2] Harald Sack[1]

Abstract: Short text categorization is an important task due to the rapid growth of online available short texts in various domains such as web search snippets, news feeds, etc. Most of the traditional methods suffer from sparsity and shortness of the text. Moreover, supervised learning methods require a significant amount of training data and manually labeling such data can be very time-consuming and costly. In this study, we propose a novel probabilistic model for Knowledge-Based Short Text Categorization (KBSTC), which does not require any labeled training data to categorize a short text [Tü].

Keywords: Short Text Categorization; Dataless Text Classification, Network Embeddings

1 Introduction

Short text categorization [Tü18b, Tü18a] plays a fundamental role in many Natural Language Processing applications such as web search, question answering, etc. Although, traditional text classification methods perform well on long text such as news articles, yet, by considering short text, most of them suffer from issues such as data sparsity and insufficient text length. Moreover, most text classification approaches require a significant amount of labeled training data and a sophisticated parameter tuning process. Manual labeling of such data can be a rather time-consuming and costly task. Especially, if the text to be labeled is of a specific scientific or technical domain, crowd-sourcing based labeling approaches do not work successfully and only expensive domain experts are able to fulfill the manual labeling task. Alternatively, semi-supervised text classification approaches have been proposed to reduce the labeling effort. Yet, due to the diversity of the documents in many applications, generating small training set for semi-supervised approaches still remains an expensive process.To address the lack of labeled data problem, we propose a novel probabilistic model for Knowledge-Based Short Text Categorization (KBSTC), which does not require any labeled training data. It is able to capture the semantic relations between the entities represented in a short text and the predefined categories by embedding both into a common vector space using the proposed network embedding technique. Finally, the appropriate category for the given text can be derived based on the semantic similarity between entities

[1] FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Germany {name.surname}@fiz-karlsruhe.de
[2] L3S Research Center, Leibniz University of Hannover, Hannover, Germany {surname}@l3s.de

doi:10.18420/inf2019_45

present in the given text and the set of predefined categories. The similarity is computed based on the vector representation of entities and categories.

2 Knowledge-Based Short Text Categorization (KBSTC)

Given an input short text t that contains a set of entities $E_t \subseteq E$ as well as a set of predefined categories $C' \subseteq C$ (from the underlying knowledge base KB), the output of the KBSTC task is the most relevant category $c_i \in C'$ for the given short text t, i.e., we compute the category function $f_{cat}(t) = c_i$, where $c_i \in C'$.
The proposed categorization task is formalized as estimating the probability of $P(c|t)$ of each predefined category c and an input text t. Based on Bayes' theorem, the probability $P(c|t)$ can be rewritten as follows:

$$P(c|t) = \frac{P(c,t)}{P(t)} \propto P(c,t) \tag{1}$$

where the denominator $P(t)$ has no impact on the ranking of the categories. Moreover, we define a novel graph embedding that exploits (a) the entity-entity graph defined via Wikilinks and (b) the entity-category graph defined by the Wikipedia category system to implement KBSTC. More details about the probability estimation can be found in [Tü].

3 Experimental Results and Conclusion

To demonstrate the performance of the KBSTC approach, it has been compared against several text classification approaches. The experimental results have proven that by utilizing KBSTC it is possible to categorize short text in an unsupervised way with a high accuracy. Further, to assess the quality of the proposed entity and category embedding model, we have compared it with state-of-the-art embedding approaches in the context of the KBSTC task. The results indicate that our embedding model enables to capture better semantic relations between entities and categories from Wikipedia. Overall, all the experimental results have demonstrated that for short text categorization, KBSTC achieves a high accuracy without requiring any labeled data, a time-consuming training phase, or a cumbersome parameter tuning step.

References

[Tü] Türker, Rima; Zhang, Lei; Koutraki, Maria; Sack, Harald: Knowledge-Based Short Text Categorization Using Entity and Category Embedding. In: ESWC 2019.

[Tü18a] Türker, Rima; Zhang, Lei; Koutraki, Maria; Sack, Harald: TECNE: Knowledge Based Text Classification Using Network Embeddings. In: EKAW. 2018.

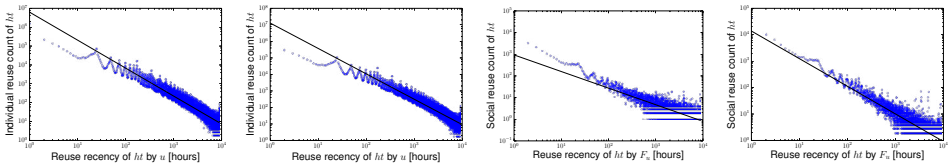
[Tü18b] Türker, Rima; Zhang, Lei; Koutraki, Maria; Sack, Harald: "the less is more" for text classification. In: SEMANTiCS. 2018.

The Impact of Time on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach

Presentation of work originally published in the Proc. of the 26th Intl. Conf. on WWW

Elisabeth Lex[1], Dominik Kowald[2]

Abstract: In our work [KPL17], we study temporal usage patterns of Twitter hashtags, and we use the Base-Level Learning (BLL) equation from the cognitive architecture ACT-R [An04] to model how a person reuses her own, individual hashtags as well as hashtags from her social network. The BLL equation accounts for the time-dependent decay of item exposure in human memory. According to BLL, the usefulness of a piece of information (e.g., a hashtag) is defined by how frequently and how recently it was used in the past, following a time-dependent decay that is best modeled with a power-law distribution. We used the BLL equation in our previous work to recommend tags in social bookmarking systems [KL16]. Here [KPL17], we adopt the BLL equation to model temporal reuse patterns of individual (i.e., reusing own hashtags) and social hashtags (i.e., reusing hashtags, which has been previously used by a followee) and to build a cognitive-inspired hashtag recommendation algorithm. We demonstrate the efficacy of our approach in two empirical social networks crawled from Twitter, i.e., *CompSci* and *Random* (for details about the datasets, see [KPL17]). Our results show that our approach can outperform current state-of-the-art hashtag recommendation approaches.



(a) Individual hashtag reuse R^2 = .883 *CompSci* dataset

(b) Individual hashtag reuse R^2 = .894 *Random* dataset

(c) Social hashtag reuse R^2 = .689 *CompSci* dataset

(d) Social hashtag reuse R^2 = .771 *Random* dataset

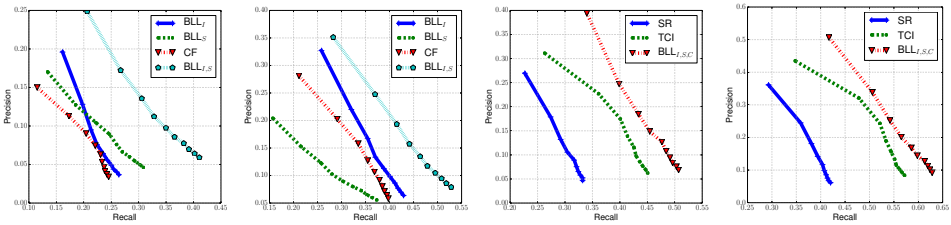
Abb. 1: The effect of time on individual and social hashtag reuse (plots are in log-log scale).

Temporal Effects of Hashtag Reuse. To determine the plausibility of our approach, we study hashtag use in our two empirical Twitter datasets, i.e., *CompSci* and *Random*. For each hashtag assignment, we investigate whether the hashtag has either been used by the same user before ("individual"), by some of her followees ("social"), by both ("individual/social"), by anyone else in the dataset ("network") or by neither of them ("external"). We find that depending on the dataset, individual or social hashtag reuse can explain approximately two-third of hashtag assignments. Also, both reuse types follow a time-dependent decay that follows a power-law distribution, as shown in Figure 1. This motivates our idea to model both individual as well as social hashtag reuse and to recommend hashtags for new tweets with the BLL equation.

[1] Graz University of Technology, ISDS, Inffeldgasse 13/V, 8010 Graz, Austria elisabeth.lex@tugraz.at
[2] Know-Center, Social Computing Group, Inffeldgasse 13/VI, 8010 Graz, Austria dkowald@know-center.at

doi:10.18420/inf2019_46

Experiments and Results. We implement the BLL equation in two variants, where the first one (i.e., $BLL_{I,S}$) predicts the hashtags of a user solely based on past hashtag usage, and the second one (i.e., $BLL_{I,S,C}$) combines $BLL_{I,S}$ with a content-based tweet analysis to also incorporate the text of the currently proposed tweet of a user. We evaluate our approach using standard evaluation protocols and metrics, and we find that our approach provides significantly higher prediction accuracy and ranking estimates than current state-of-the-art hashtag recommendation algorithms in both scenarios (for more details about the baselines, refer to [KPL17]), as shown in Figure 2.



(a) *Scenario 1*: Hashtag rec. w/o current tweet *CompSci* dataset (b) *Scenario 1*: Hashtag rec. w/o current tweet *Random* dataset (c) *Scenario 2*: Hashtag rec. w/ current tweet *CompSci* dataset (d) *Scenario 2*: Hashtag rec. w/ current tweet *Random* dataset

Abb. 2: Precision / Recall plots of our evaluation scenarios for k = 1 - 10 recommended hashtags.

Conclusion and Reproducibility. We find that temporal effects play an important role in hashtag reuse on Twitter. We propose our cognitive-inspired hashtag recommendation approaches $BLL_{I,S}$ and $BLL_{I,S,C}$ to account for such effects. We compare both algorithms to state-of-the-art hashtag recommendation algorithms and find that our cognitive-inspired approaches outperform these algorithms in terms of prediction accuracy and ranking. With our work, we aim to contribute to the rich line of research on improving the use of hashtags in social networks. We also hope to spark future work to utilize models from human memory theory to model and explain digital traces and user behavior online. For the sake of reproducibility, we implement and evaluate our approach by extending our open-source tag recommender benchmarking framework *TagRec*. The source code and framework are freely accessible for scientific purposes on the Web[3].

Keywords: hashtag recommendation, ACT-R, temporal effects, hashtag reuse, user behavior modeling

Literaturverzeichnis

[An04] Anderson, John R; Bothell, Daniel; Byrne, Michael D; Douglass, Scott; Lebiere, Christian; Qin, Yulin: An integrated theory of the mind. Psychological review, 111(4):1036, 2004.

[KL16] Kowald, Dominik; Lex, Elisabeth: The Influence of Frequency, Recency and Semantic Context on the Reuse of Tags in Social Tagging Systems. In: Proceeding of Hypertext'16. ACM, S. 237–242, 2016.

[KPL17] Kowald, Dominik; Pujari, Subhash Chandra; Lex, Elisabeth: Temporal Effects on Hashtag Reuse in Twitter: A Cognitive-Inspired Hashtag Recommendation Approach. In: Proceedings of the 26th International Conference on World Wide Web. 2017.

[3] https://github.com/learning-layers/TagRec

What If We Encoded Words as Matrices and Used Matrix Multiplication as Composition Function?

Presentation of work originally published in the Proc. of the International Conference on Learning Representations 2019

Lukas Galke[1] Florian Mai[2] Ansgar Scherp[3]

Abstract: We summarize our contribution to the International Conference on Learning Representations *CBOW Is Not All You Need: Combining CBOW with the Compositional Matrix Space Model*, 2019. We construct a text encoder that learns matrix representations of words from unlabeled text, while using matrix multiplication as composition function. We show that our text encoder outperforms continuous bag-of-word representations on 9 out of 10 linguistic probing tasks and argue that the learned representations are complementary to the ones of vector-based approaches. Hence, we construct a hybrid model that jointly learns a matrix and a vector for each word. This hybrid model yields higher scores than purely vector-based approaches on 10 out of 16 downstream tasks in a controlled experiment with the same capacity and training data. Across all 16 tasks, the hybrid model achieves an average improvement of 1.2%. These results are insofar promising, as they open up new opportunities to efficiently incorporate order awareness into word embedding models.

Keywords: machine learning; natural language processing; representation learning

Introduction Word embeddings [CW08, Mi13] are celebrated as one of the most impactful contributions from unsupervised representation learning to natural language processing [Go16]. After unsupervised learning from a large textual corpus, the word embeddings can be transferred to various downstream tasks. Sentence representations are then composed of the sum or the mean of the words in the sentence, the so-called continuous bag-of-words [Mi13]. Since these operations are inherently commutative, any information of word order is lost. For instance, the following two sentences would yield the exact same embedding: "The movie was not awful, it was rather great." and "The movie was not great, it was rather awful." A classifier based on the continuous bag-of-words embedding of these sentences would inevitably fail to distinguish the two different meanings [Go17, p. 151]. While using n-grams is a common choice to bring order-awareness into traditional classifiers, storing embeddings for all n-gram combinations would require exponential space. Other approaches such as contextualized word representations [Pe18] require substantially more parameters. We identify the need for efficient, order-aware, word embedding models.

[1] Kiel University / ZBW, Germany lga@informatik.uni-kiel.de
[2] Idiap Research Institute, Martigny, Switzerland florian.mai@idiap.ch
[3] University of Essex, United Kingdom ansgar.scherp@essex.ac.uk

doi:10.18420/inf2019_47

Approach We propose to encode each word as a matrix and to use matrix multiplication as composition function. Because of the associative property, merely $O(\log n)$ sequential steps are sufficient to encode a sentence. Frequent n-grams can be precomputed via dynamic programming. The idea was theoretically explored earlier by Rudolph and Giesbrecht [RG10] as the compositional matrix space model of language without providing any learning algorithm. We show that the CBOW training objective [Mi13] can be adapted to obtain an unsupervised and efficient training scheme by making two adaptions: On the one hand, we modify the initialization scheme such that the expected value of chained matrix multiplications is constant. On the other hand, we chose a random word as target instead of the center word to alleviate bias.

Results and Conclusion Our experiments [MGS19] show that matrix-based embeddings yield an increase in 9 out of 10 linguistic probing tasks compared to vector-based embeddings. We find that matrix-based and vector-based models complement each other well. When training a joint model with both matrix- and a vector-based components, the model yields an increased performance on 10 out of 16 downstream tasks compared the vector-based approach trained on the same data with the same capacity. The average improvement across all 16 tasks is 1.2%. These results are insofar promising, as they open up new opportunities to efficiently incorporate order awareness into word embedding models.

Acknowledgement This research was supported by the Swiss National Science Foundation under grant number "FNS-30216".

References

[CW08] Collobert, Ronan; Weston, Jason: A unified architecture for natural language processing: deep neural networks with multitask learning. In: ICML. ACM, 2008.

[Go16] Goth, Gregory: Deep or shallow, NLP is breaking out. Commun. ACM, 59(3), 2016.

[Go17] Goldberg, Yoav: Neural Network Methods for Natural Language Processing. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2017.

[MGS19] Mai, Florian; Galke, Lukas; Scherp, Ansgar: CBOW Is Not All You Need: Combining CBOW with the Compositional Matrix Space Model. In: International Conference on Learning Representations. 2019.

[Mi13] Mikolov, Tomas; Sutskever, Ilya; Chen, Kai; Corrado, Gregory S.; Dean, Jeffrey: Distributed Representations of Words and Phrases and their Compositionality. In: NIPS. 2013.

[Pe18] Peters, Matthew E.; Neumann, Mark; Iyyer, Mohit; Gardner, Matt; Clark, Christopher; Lee, Kenton; Zettlemoyer, Luke: Deep Contextualized Word Representations. In: NAACL-HLT. Association for Computational Linguistics, 2018.

[RG10] Rudolph, Sebastian; Giesbrecht, Eugenie: Compositional Matrix-Space Models of Language. In: ACL. The Association for Computer Linguistics, 2010.

Debiasing Vandalism Detection Models at Wikidata (Extended Abstract)

Presentation of work originally published in WWW '19 [He19]

Stefan Heindorf[1] Yan Scholten[2] Gregor Engels[3] Martin Potthast[4]

The ethics of artificial intelligence have become a major societal issue, evidenced also by them becoming a focus of attention of policy making. Recently, the European Union released "Ethics Guidelines for Trustworthy AI" — as did the IEEE and major companies such as Google, Microsoft, and IBM.[5] A central point in all the guidelines is the fairness of machine learning models and the mitigation of discrimination against minorities based on biased models. In this extended abstract, we report on a case study on debiasing vandalism detection models at Wikidata, the crowdsourced knowledge base of the Wikimedia Foundation.

Knowledge bases play an important role in modern information systems. For instance, web search engines use them to enrich search results, conversational agents to answer factual questions, and fake news detectors for fact-checking. Collecting knowledge at scale still heavily relies on crowdsourcing: Google acquired the open Freebase project to bootstrap its proprietary "Knowledge Graph" until Freebase was shut down and succeeded by Wikidata, the free knowledge base of Wikimedia. Other prominent open knowledge bases like YAGO and DBpedia also depend on crowdsourcing by extracting knowledge from Wikipedia. As crowdsourcing knowledge has a long history, so does the fight against damage caused by vandals and other users, which may propagate to information systems using the knowledge base, potentially reaching a wide audience.

Wikidata makes for an interesting case study to analyze and mitigate biases as it has one of the largest online communities and provides opportunities to pay particular attention to the content rather than the user reputation. Unfortunately, it is still common practice to identify malicious edits via meta data such as geolocation of IP addresses, age of user account, or language of edited content. While those features are simple to obtain, they do not directly judge the quality of an edit and harm well-intentioned users.

[1] Paderborn University, heindorf@uni-paderborn.de
[2] Paderborn University, yascho@mail.uni-paderborn.de
[3] Paderborn University, engels@uni-paderborn.de
[4] Leipzig University, martin.potthast@uni-leipzig.de
[5] https://ec.europa.eu/futurium/en/ai-alliance-consultation/; https://ethicsinaction.ieee.org/; https://ai.google/principles/; https://www.microsoft.com/en-us/ai/our-approach-to-ai; https://www.ibm.com/watson/ai-ethics/

doi:10.18420/inf2019_48

In our research, we revealed for the first time that state-of-the-art vandalism detectors employed at Wikidata [He16, SHT17] are heavily biased against certain groups of editors. For example, *benign* edits of anonymous users receive vandalism scores over 300 times higher than *benign* edits of registered users. Such a widespread discrimination of certain user groups (especially that of anonymous editors) undermines the founding principles on which Wikimedia's projects are built.[6] Although the discrimination of anonymous users has long been recognized and the problem has been tackled through community outreach,[7] when discrimination gets encoded into automatic decision-making, this aggravates the problem. For example, it has been previously found that desirable newcomers whose edits are automatically reverted are much more likely to withdraw from the project [Ha13].

We carefully analyzed different sources of bias in Wikidata's damage control system and developed two new machine learning models that significantly reduce bias compared to the state-of-the-art. Our model FAIR-E uses graph embeddings to check the content's correctness without relying on biased user features. Our model FAIR-S selects the best-performing hand-engineered features under the constraint of no user features. Furthermore, we experiment with different transformations of the state-of-the-art vandalism detector WDVD: post-processing scores, reweighting training samples, and combining approaches via ensembles. We evaluate our approaches on a subset of the standardized, large-scale Wikidata Vandalism Detection Corpus 2016 [He17], and compare our results to others from the literature. Our best model FAIR-S reduces the bias ratio of WDVD from 310.7 to only 11.9, while maintaining high predictive performance at 0.963 ROC_{AUC} and 0.316 PR_{AUC}.

Keywords: Bias; Fairness; Knowledge Base; Wikidata; Data Quality; Vandalism

Acknowledgements: This work was supported by the German Research Foundation within the Collaborative Research Center "On-The-Fly Computing" (CRC 901).

Bibliography

[Ha13] Halfaker, A.; Geiger, R. S.; Morgan, J. T.; Riedl, J.: The Rise and Decline of an Open Collaboration System: How Wikipedia's Reaction to Popularity is Causing Its Decline. American Behavioral Scientist, 57(5), 2013.

[He16] Heindorf, S.; Potthast, M.; Stein, B.; Engels, G.: Vandalism Detection in Wikidata. In: CIKM. 2016.

[He17] Heindorf, S.; Potthast, M.; Engels, G.; Stein, B.: Overview of the Wikidata Vandalism Detection Task at the WSDM Cup 2017. In: WSDM Cup. 2017.

[He19] Heindorf, S.; Scholten, Y.; Engels, G.; Potthast, M.: Debiasing Vandalism Detection Models at Wikidata. In: WWW. 2019.

[SHT17] Sarabadani, A.; Halfaker, A.; Taraborelli, D.: Building Automated Vandalism Detection Tools for Wikidata. In: WWW (Companion Volume). 2017.

[6] https://meta.wikimedia.org/wiki/Founding_principles

[7] https://en.wikipedia.org/wiki/Wikipedia:IPs_are_human_too

Track 4 - Informatik mit Recht

Informatik mit Recht

Rüdiger Grimm,[1] Gerrit Hornung,[2] Christoph Sorge,[3] Indra Spiecker gen. Döhmann,[4] Arno Wacker[5]

Informatik und Recht stellen sich gegenseitig vor neue Herausforderungen: Informatische Innovationen können etablierte Regelungskonzepte unterlaufen, Gesetze wie die Datenschutz-Grundverordnung neuartige Vorgaben für Design und Implementierung schaffen. Der Track beleuchtete dieses Spannungsfeld mit einem Fokus auf Datenschutz und Datensicherheit und adressierte die interdisziplinären Herausforderungen, denen sich Informatikerinnen und Informatiker einerseits, Juristinnen und Juristen andererseits im Dienste der Gesellschaft stellen müssen. Innerhalb des Tracks wurde die Tradition des auf den letzten GI-Jahrestagungen etablierten Workshops „Recht und Technik: Datenschutz im Diskurs“ fortgeführt.

Mögliche Themen umfassten: aktuelle Herausforderungen für die IT-Sicherheit; neue Anforderungen der DSGVO an die IT-Sicherheit und technische Mechanismen zur Umsetzung, insbesondere System- und Selbstdatenschutz; technische Aspekte der IT-Sicherheit (wie z.B. Privacy Enhancing Mechanisms, Penetration Testing, Zugriffskontrollen); Umsetzungsstrategien für den Datenschutz in Organisationen; Datenschutz-Compliance; Meldepflichten bei IT-Sicherheitsvorfällen; Regulierung der IT-Sicherheit in Deutschland in Europa; weltweite Vernetzung und weltweite Regulierung von Datenschutz und IT-Sicherheit; ökonomische Aspekte der IT-Sicherheit; IT-Sicherheit aus Nutzerperspektive; Rechtsfragen von datenbasierten Geschäftsmodellen und „Dateneigentum“; IT-Sicherheit im Internet of Things: Industrie 4.0 und Smart Home; Künstliche Intelligenz – Fähigkeiten, Entlastung, Verantwortung; automatisiertes Entscheiden: Diskriminierung, Folgen, Kontrolle; E-Government und E-Justice.

Das Programmkomitee bestand aus

- Matthias Bäcker, Universität Mainz
- Jens-Matthias Bohli, Hochschule Mannheim
- Thomas Bräuchle, Rechtsanwalt, Stuttgart
- Katharina Bräunlich, Universität Koblenz-Landau
- Felix Drefs, Heuking Kühn Lüer Wojtek

[1] Universität Koblenz-Landau, Institut für Wirtschafts- und Verwaltungsinformatik, grimm@uni-koblenz.de
[2] Universität Kassel, Institut für Wirtschaftsrecht, gerrit.hornung@uni-kassel.de
[3] Universität des Saarlandes, Professur für Rechtsinformatik, christoph.sorge@uni-saarland.de
[4] Goethe Universität Frankfurt a.M., Forschungsstelle Datenschutz, spiecker@jur.uni-frankfurt.de
[5] Universität der Bundeswehr München, Datenschutz und Compliance, arno.wacker@unibw.de

doi:10.18420/inf2019_49

- Matthias Enzmann, Frauhofer SIT, Darmstadt
- Rüdiger Grimm, Universität Koblenz (Co-Chair)
- Nils Gruschka, Universität Oslo
- Christoph Gusy, Universität Bielefeld
- Marit Hansen, ULD Schleswig-Holstein
- Niko Härting, Rechtsanwalt, Berlin
- Gerrit Hornung, Universität Kassel (Co-Chair)
- Walter Hötzendorfer, Universität Wien
- Thomas Kahler, Datenschutzbeauftragter
- Thomas Kehr, Dornbach GmbH Rechtsanwaltsgesellschaft
- Thomas Lapp, Rechtsanwalt, IT-Kanzlei Dr. Lapp
- Kai von Lewinski, Universität Passau
- Mario Lischka, DZ Bank
- Ronald Petrlic, LfDI Baden-Württemberg
- Burkhard Schafer, Universität Edinburgh
- Christoph Sorge, Universität des Saarlandes (Co-Chair)
- Indra Spiecker gen. Döhmann, Goethe-Universität Frankfurt a.M. (Co-Chair)
- Juergen Taeger, Universität Oldenburg
- Arno Wacker, Universität der Bundeswehr München (Co-Chair)

Eingereicht wurden 21 Beiträge mit einer großen Bandbreite von Themen, von denen 12 für den Track berücksichtigt wurden. Für die vier Sessions wurden diese in vier thematische Bereiche gegliedert. Diese befassten sich mit

- Der Unterstützung der Umsetzung der DSGVO (drei Beiträge)
- Dem rechtsgemäßen Umgang mit Daten (drei Beiträge)
- Der digitalen Forensik und Ermittlung (zwei Beiträge)
- Rechtskonformen Anwendungen (vier Beiträge)

Mit den eingereichten und ausgewählten Beiträgen und den inhaltlichen Strukturierungen führte der Workshop innovative Ansätze in interdisziplinärer Weise zur Bewältigung technischer und rechtlicher Problemlagen zusammen.

Full Papers

Muster zur praxisorientierten Umsetzung und konformen Nutzung der DSGVO

Daniel Rösch[1], Thomas Schuster[1], Lukas Waidelich[1] und Sascha Alpers[2], Wasilij Beskorovajnov[2], Roland Gröll[2] und Hoa Tran[2]

Abstract: In der heutigen Wissensgesellschaft ist der Austausch von sensiblen Informationen von essenzieller Bedeutung. Sowohl unternehmensbezogene als auch persönliche Daten werden durch Nachrichten (z.B. E-Mail) oder Freigaben über Cloud-Dienste (z.B. OwnCloud) ausgetauscht. Neben individuellen Interessen unterliegt dieser Datenaustausch gesetzlichen Regularien. Seit Mai 2018 ist die Europäische Datenschutz-Grundverordnung (DSGVO) voll wirksam. Die Regularien und die möglichen Strafen führen derzeit noch in vielen Organisationen zu Unsicherheiten. Dieser Artikel zeigt Möglichkeiten auf, wie die tägliche Arbeit konform zur DSGVO gestaltet werden kann. Hierzu definieren wir Muster, welche die bestehenden Anforderungen der DSGVO in technische Lösungsansätze zur Umsetzung konformer Informationsdienste überführen. Der Artikel geht dabei beispielhaft auf einige DSGVO-Anforderungen ein. Durch die Muster wollen wir die Unsicherheiten reduzieren und die Umsetzung der DSGVO erleichtern. Die beschriebenen Muster können als Referenzkatalog für die Anbieter und Nutzer von Informationsdiensten dienen. Zur Demonstration der praktischen Umsetzung nutzen wir beispielhaft ein Anwendungssystem aus dem Forschungsprojekt Einfaches Digitales Vergessen (EDV).

Keywords: DSGVO, Datenschutz Muster, Handlungsempfehlungen, Muster zu Rechten der betroffenen Person, Muster zu Pflichten des Verantwortlichen, Cloud-Dienste.

1 Einleitung

In den vergangenen Jahren wurden viele Technologien entwickelt, die in besonderem Maß die Erzeugung und Verarbeitung großer Datenvolumina vorantreiben. Eines dieser Trendthemen im Kontext des digitalen Wandels ist Internet of Things (IoT). Durch IoT werden enorme Mengen an Daten erzeugt, die bei der Interaktion analoger und digitaler Welt beinahe kontinuierlich anfallen. Weitere technologisch orientierte Ansätze wie Big Data zielen auf die effiziente Verarbeitung dieser Datenmengen ab. Der technologische Wandel wird inzwischen auch in der Gesellschaft wahrgenommen. Besonders die Auswertung großer Datenmengen und die Analyse persönlicher Verhaltensweisen wird zunehmend kritisch betrachtet. Vor diesem Hintergrund wurde auf europäischer Ebene die neue Europäische Datenschutz-Grundverordnung (DSGVO) beschlossen, welche nach Ablauf einer Übergangsfrist seit dem 25. Mai 2018 zur Anwendung kommt. Die Harmonisierung des Datenschutzes hin zu einem Verbot mit Erlaubnisvorbehalt und die Wir-

[1] Hochschule Pforzheim, Tiefenbronner Straße 65, 75175 Pforzheim,
{daniel.roesch, thomas.schuster, lukas.waidelich}@hs-pforzheim.de
[2] FZI Forschungszentrum Informatik, Haid- und Neu-Straße 10-14, 76131 Karlsruhe, {alpers, beskorovajnov, groell, tran}@fzi.de

doi:10.18420/inf2019_50

kung der DSGVO auf Daten europäischer Bürger bei ausländischen, hier angebotenen Diensten verbessern den Schutz von natürlichen Personen gegenüber den Gefahren der Verarbeitung personenbezogener Daten [Sc18]. Mit dem Artikel wollen wir deutlich machen, dass die DSGVO vielmehr eine Vorgabe ist, die zu einer zeitgemäßen Technologieentwicklung und -nutzung beitragen kann. Dabei sollen auch die Rechte der EU-Bürger hinsichtlich ihrer digitalen Datensouveränität [BMR18] schon bei der ethisch verantwortungsvollen Entwicklung von Technologien berücksichtigt werden. Die technische Herausforderung ist die Erfüllung der DSGVO durch standardisierte Technologien. Dabei muss der Verantwortliche geeignete technische und organisatorische Maßnahmen treffen, um die Rechte der betroffenen Person zu schützen.

Im Kern soll dieses Papier das Verständnis für die Anforderungen der DSGVO im Kontext von technischen Umsetzungen verbessern. Hierfür werden Muster definiert, welche die Anforderungen beschreiben und technische Lösungsstrategien zur konformen Umsetzung liefern. Um die Wiederverwendbarkeit von technischen Lösungsansätzen zu fördern, werden auch Querverbindungen zwischen den Anforderungen aufgezeigt – hierzu werden die Muster entsprechend miteinander verknüpft. Die Muster sind nach einem definierten Schema (Problemstellung, Kontext, Lösungsansatz) aufgebaut, wie dies aus anderen Forschungsgebieten [Al95] und besonders der Softwaretechnik bekannt ist [Fo03; Ga08; HW04]. Die Muster können so als Blaupausen verstanden werden, die Lösungsstrategien für Fragen anbieten, welche in der täglichen Arbeit auftreten können. Der Entwurf der Muster folgt dem konstruktivistischen Paradigma der Designwissenschaft. Neues Wissen wird gewonnen indem Artefakte in Form von Modellen, Methoden oder Systemen beschrieben werden. Im Gegensatz zu empirischen Forschung ist das Ziel nicht unbedingt, die Gültigkeit von Forschungsergebnissen im Hinblick auf ihre Wahrheit zu bewerten, sondern den Nutzen als Werkzeug zur Lösung bestimmter Probleme darzustellen [He04].

Vor dem Hintergrund der beschriebenen Ausgangsituation, soll in diesem Artikel die folgende Fragestellung beantwortet werden: Können aus der DSGVO technische Anforderungen und passende Lösungsansätze in Form von Mustern abgeleitet werden? Aus dieser primären Fragestellung leiten sich folgende untergeordnete Forschungsfragen ab:

Q1. Lassen sich die Anforderungen aus der DSGVO in Mustern zusammenfassen?

Q2. Können auf Basis der identifizierten Muster DSGVO konforme technische Lösungen konzipiert und umgesetzt werden?

2 Grundsätze zur Muster-Entwicklung im Kontext der DSGVO

In der DSGVO werden mehrere Grundsätze für die Verarbeitung von personenbezogenen Daten beschrieben. Diese sind u.a. *Rechtmäßigkeit, Verarbeitung nach Treu und Glauben, Transparenz und Nachvollziehbarkeit*, *Zweckbindung*, *Datenminimierung*, *Richtigkeit*, *Speicherbegrenzung*, *Integrität und Vertraulichkeit und Rechenschaftspflicht* (Artikel 5).

Manche der Grundsätze lassen sich nicht vollumfänglich durch technische Maßnahmen erfüllen oder die Umsetzung weist einen sehr hohen Aufwand auf. Dazu gehören: *Rechtmäßigkeit nach Treu und Glauben, Zweckbindung, Integrität und Vertraulichkeit sowie Rechenschaftspflicht*. Die Betrachtung dieser Grundsätze muss primär auf anderen Ebenen (bspw. Geschäftsprozesse, Geschäftsmodelle) erfolgen. Erst nach deren Betrachtung können dann technische Konsequenzen abgeleitet werden [APOR18].

In diesem Artikel sollen daher nur die Grundsätze, die mit vertretbarem Aufwand (unmittelbar) durch technische Vorkehrungen umgesetzt werden können, betrachtet werden. Somit können besonders die folgenden DSGVO-Grundsätze hier in Mustern zusammengefasst werden: *Transparenz und Nachvollziehbarkeit, Zweckbindung, Datenminimierung und Speicherbegrenzung*. Durch die Formulierung passender Muster, kann die Forschungsfrage Q1 bereits erfolgreich beantwortet werden. Zur DSGVO sind bisher keine Muster und Lösungen veröffentlicht worden. Lediglich Huth konkludiert in seinem Artikel [Hu17] wie sich die neuen Regularien der DSGVO auf bestehende Unternehmen, deren Prozesse und Systeme auswirken. Der Artikel greift bestehende Lösungen auf, erweitert diese durch Lösungsansätze und leitet daraus generalisierte Musterbeschreibungen ab.

Die Methodik zur Beschreibung der Muster weist stets eine einheitliche Struktur auf. Wir orientieren unser Vorgehen an bekannten Veröffentlichungen zu Mustern aus anderen Bereichen. Um die Muster für unsere Problemstellungen und Lösungen optimal darzustellen zu können, haben wir einige Strukturanpassungen in der Musterbeschreibung vorgenommen. Dementsprechend verstehen wir ein Muster als eine Blaupause, die für eine gegebene Problemstellung (und einen definierten Kontext) einen generalisierten Lösungsansatz (Strategie) bereitstellt. Dieses Paradigma ist in vielen Disziplinen bekannt und wurde zunächst in der Architektur und später in der Software-Entwicklung populär [Al95; Be02; BH04; Bu04; Fo03; Ga08; Wi12]. Häufig werden themenorientiert ganze Kataloge an Mustern beschrieben. Zahlreiche Publikationen beschreiben weitere Ansätze zur Entwicklung von Mustern [Ro06; Sc03; Sc04; SNL05; Wi12]. Im Bereich Security wurden bereits einige Muster identifiziert und beschrieben (siehe auch [BH04; SNL05]). Wie bereits oben erwähnt, sind im Bereich der DSGVO bislang keine Muster vorhanden. Im Forschungsbereich Datenschutz sind insgesamt wenige Musterbeschreibungen vorhanden. [Ro06] beschreibt Datenschutz-Muster in Hinblick auf Online-Transaktionen. [Ah07] definieren Datenschutzmuster und -überlegungen im Bereich der Online- und mobilen Fotoübertragung.

3 Entwicklung und Analyse von Datenschutzmustern

Insgesamt können die entwickelten Muster in drei inhaltliche Schwerpunkte untergliedert werden. Dazu gehören allgemeine Muster (I), Rechte der betroffenen Person (II) und Pflichten der Verantwortlichen (III). Jedes unserer Muster wird in einem separaten Teilkapitel innerhalb dieser Schwerpunkte vorgestellt und mit einer eindeutigen Muster-

bezeichnung identifiziert. Somit ist eine schnelle und gezielte Auffindbarkeit eines Lösungsmusters zu einem von der DSGVO auferlegten Prinzip gegeben. Dabei werden in dieser Veröffentlichung nur besonders relevante Prinzipien adressiert, weitere Prinzipien können und sollten später nach dem gleichen Schema ergänzt werden. Im ersten Abschnitt jedes Musters werden Anforderungen, welche die DSGVO fordert näher beschrieben und auf den Verordnungstext referenziert. Im Absatz *Resultierende Herausforderung* wird die daraus resultierende Problemstellung genauer beleuchtet. Es erfolgt eine klare Identifikation des Problems sowie der betroffenen Komponenten. Im dritten Teil *Technischer Lösungsansatz* werden verschiedene Wege zur Lösung der Problemstellung präsentiert. Bei der Umsetzung muss dann jeweils entschieden werden welcher Lösungsweg für den jeweiligen Kontext am besten umsetzbar ist. Abschließend wird eine Checkliste beschrieben, mit der der Anwender prüfen kann, ob alle DSGVO-Anforderungen erfüllt wurden. In jeder Kategorie werden nachfolgend nur einige, wichtige Muster ausführlich erklärt.

3.1 Allgemeine Muster (I)

Dieses Kapitel beschreibt zentrale Anforderungen aus der DSGVO, die als Muster zusammengefasst sind. Insgesamt wurden in diesem Bereich fünf Muster identifiziert: *Transparenz und Nachvollziehbarkeit (1), Zweckbindung (2), Datenminimierung (3), Richtigkeit (4) und Speicherbegrenzung (5)*. Die folgenden Abschnitte beschreiben die aus unserer Sicht wichtigsten Muster.

Transparenz und Nachvollziehbarkeit (1)

Vorgabe der DSGVO: Personenbezogene Daten müssen in einer für die betroffene Person nachvollziehbaren Weise verarbeitet werden. Art. 5 Abs. 1a.

Resultierende Herausforderung: Die Art und Weise sowie alle relevanten Daten, die im Rahmen eines Dienstes in Bezug auf eine Person verarbeitet werden, sind auszuweisen und offenzulegen. Die Offenlegung und Ausweisung sind kontinuierliche Anforderungen an den Dienst. Die Kern-Herausforderung liegt in der Bereitstellung einer Schnittstelle, die der betroffenen Person alle, für die Transparenz und Nachvollziehbarkeit, notwendigen Informationen nach Bedarf offenlegt.

Technischer Lösungsansatz: Drei technische Teilaspekte sind zu beachten, um der Transparenz und Nachvollziehbarkeit nachzukommen.

1. *Übersicht erhebender Daten*: Schon vor der Benutzung des Dienstes muss der Anbieter eine Liste mit allen Daten, die erhoben werden sollen, vorlegen. Hierzu kann der technische Lösungsansatz der *Informationspflicht* verwendet werden.
2. *Offenlegung der gespeicherten Daten:* Hierfür empfiehlt sich die technische Maßnahme, um den Nutzern das Recht auf Auskunft zu ermöglichen.

3. *Offenlegung der Art und Weise der Verarbeitung:* Dieser Teilaspekt stellt eine besondere Herausforderung dar. Idealerweise werden alle relevanten Prozesse für die betroffene Person in transparenter Art und Weise offengelegt. Die Offenlegung von Code kann, zusätzlich zur Datenschutzerklärung, technisch versierten Personen ein tiefergehendes Verständnis der Art und Weise der Verarbeitung ermöglichen. Mindestens jedoch sollte der Anbieter vor der Erhebung der Daten die Leitfragen aus der folgenden Checkliste beantworten.

Checkliste:
- Wird der Verarbeitungsprozess von personenbezogenen Daten erläutert?
- Sind die Fragen aus der Checkliste des Musters *Informationspflicht (6)* beantwortet?
- Ist die potenzielle Datenweitergabe in der Erklärung ausführlich beschrieben?

Zweckbindung (2)

Vorgabe der DSGVO: Personenbezogene Daten dürfen nur für festgelegte, eindeutige und legitime Zwecke erhoben werden und dürfen nicht in einer mit diesen Zwecken nicht zu vereinbarenden Weise weiterverarbeitet werden. Artikel 5 Abs. 1b.Davon gibt es nur wenige Ausnahmen wie bspw. Weiterverarbeitung für Archivzwecke im öffentlichen Interesse. Art. 89 Abs. 1.

Resultierende Herausforderung: Die Verarbeitungszwecke müssen aus der Datenschutzerklärung eindeutig erkennbar sein. Daten dürfen nur für die Verarbeitungsprozesse zugreifbar sein, die für einen angegebenen Zweck notwendig sind.

Technischer Lösungsansatz: Bereitstellung einer Erklärung, die die Zwecke der Verarbeitung von personenbezogenen Daten beschreibt. Je nach gewählter Art der Datenverarbeitung, können sind zwei Fälle unterschieden werden:

1. *Die Daten werden zentral abgelegt:* Hier empfiehlt sich eine logische Aufteilung von Verarbeitungsprozessen auf Verarbeitungszwecke. Die Ablage der Daten erfolgt zusammen mit dem deklarierten Zweck. Das ermöglicht es Verarbeitungsprozesse und abgelegte Daten zu organisieren, indem der Verarbeitungszweck als Zugriffsrichtlinie (Policy) für Prozesse dient.
2. *Die Daten werden direkt in den einzelnen Verarbeitungsprozessen gespeichert:* Dies ist eine einfache Variante, die Daten zweckgebunden zu speichern. In der Folge jedoch müssen viele Daten redundant gespeichert werden (eine Herausforderung bspw. Bei Auskunfts- und Löschansprüchen). Die Verarbeitungsprozesse müssen dennoch zu Verarbeitungszwecken zugeordnet sein.

Checkliste:
- Sind eindeutige und legitime Verarbeitungszwecke festgelegt?
- Beschreibt die Datenschutzerklärung alle Verarbeitungszwecke?

- Sind Verarbeitungsprozesse auf Verarbeitungszwecke abgebildet?
- Werden gespeicherte Daten explizit mit einem Attribut Zweck versehen oder ausschließlich zweckgebunden in isolierten Verarbeitungsprozessen gespeichert?

Datenminimierung (3)

Vorgabe der DSGVO: Personenbezogene Daten müssen dem Zweck angemessen und auf das für den Verarbeitungszweck notwendige Maß beschränkt sein. Artikel 5 Abs. 1c.

Resultierende Herausforderung: Verringerung der Menge der verarbeiteten Daten und Anzahl der Betroffenen. Des Weiteren ist die Mindestmenge an personenbezogenen Daten zu identifizieren, welche für die Verarbeitungszwecke notwendig sind.

Technischer Lösungsansatz: Bereits in der Entwurfsphase eines Softwaresystems muss das Datenmodell im Hinblick auf den Verarbeitungszweck überprüft und angepasst werden. Auch während des Betriebs können sich Änderungen bezüglich der Notwendigkeit bestimmter Daten und hinsichtlich mancher Zwecke ergeben. Dies erfordert eine Architektur, in der das Datenmodell dynamisch anpassbar ist.

Checkliste:
- Welche Datenstruktur ist gleichzeitig minimal und zweckdienlich, um den gewünschten Dienst zu erbringen?
- Wurden Daten (oder Attribute) die nicht (mehr) für Verarbeitungszwecke notwendig sind gelöscht?

Speicherbegrenzung (5)

Vorgabe der DSGVO: Personenbezogene Daten müssen so gespeichert werden, die die Identifizierung der betroffenen Personen nur so lange ermöglicht, wie es für die Verarbeitungszwecke erforderlich ist. Artikel 5 Abs. 1e.

Resultierende Herausforderung: Die Dauer der Speicherung von personenbezogenen Daten muss definiert sein. Mit der Zweckerfüllung sind personenbezogene Daten aus dem System zu entfernen oder die Verbindung zu den personenbezogenen Daten so aufzuheben, dass die Identifizierung der betroffenen Person nicht mehr möglich ist. Letzteres ist in vielen Fällen nur schwer zu erreichen ist, da betroffene Personen oftmals rückwirkend durch noch vorhandene Attribute identifiziert werden können [Sw02].

Technischer Lösungsansatz: Das Datenmodell muss einen Lebenszyklus für die Daten vorsehen. Als Grundlage bieten sich Zeit- und Zweckattribute an. Werden die Daten verschlüsselt ist eine unumkehrbare Löschung des Schlüssels ausreichend, um die Daten nicht-identifizierbar zu machen [BH04]. Andere Anonymisierungsmechanismen, wie

Differential Privacy [Dw08], sind auch möglich, aber in der Praxis nur schwer umzusetzen.

Checkliste:

- Sind die Daten mit einer Speicherdauer versehen, die aufgrund eines bestimmten Zwecks beschränkt ist?
- Falls Anonymisierung gewünscht: Ist der Speichermechanismus für die gespeicherte Art von Daten sinnvoll?
- Sofern die Daten verschlüsselt vorliegen: Kann der Schlüssel unumkehrbar gelöscht werden?

3.2 Rechte der betroffenen Person (II)

Neben allgemeinen Anforderungen aus der DSGVO werden in diesem Abschnitt alle Rechte der betroffenen Person aufgelistet und näher beschrieben, die sich aus der DSGVO ergeben. Dazu gehören die *Informationspflicht (6), das Recht auf Auskunft (7), das Recht auf Berichtigung (8), das Recht auf Löschung (9), das Recht auf Einschränkung der Verarbeitung (10) sowie das Recht auf Datenübertragbarkeit (11).* Analog zum vorhergegangenen Kapitel werden nun die wichtigsten dieser Muster erklärt. Wie im vorherigen Abschnitt werden die wichtigsten Muster veranschaulicht und erläutert.

Informationspflicht (6)

Vorgabe der DSGVO: Zum Zeitpunkt der Erhebung von personenbezogenen Daten müssen sämtliche Informationen über die Datenverarbeitung der betroffenen Person mitgeteilt werden. Artikel 13 Abs. 1, 2.

Resultierende Herausforderung: Gemäß den DSGVO-Vorgaben sind die Informationen verständlich, leicht zugänglich und in einer klaren Sprache in einer Erklärung schriftlich oder elektronisch an die betroffene Person zu übermitteln. Die Kenntnisnahme der Datenschutzerklärung ist eine zwingende Voraussetzung für eine Nutzung des Dienstes. Die Datenschutzerklärung muss außerdem immer (auch nach der Kenntnisnahme) leicht auffindbar sein (durch max. 2 Klicks).

Technischer Lösungsansatz: Die Datenschutzerklärung ist vor der Registrierung des Nutzers in der Anwendung als Text an die betroffene Person auszuweisen. Die anschließende Registrierung darf erst nach der Protokollierung der erfolgreichen Kenntnisnahme der Datenschutzerklärung möglich sein. Während der Nutzung muss die Datenschutzerklärung jederzeit leicht auffindbar sein.

Checkliste:

- Beinhaltet die Benachrichtigung nachfolgende Informationen? Name, Kontaktdaten des Verantwortlichen für die Datenerhebung, ggf. Kontaktdaten des Datenschutzbe-

auftragten, Zwecke der Datenverarbeitung und deren Rechtsgrundlage, Empfänger der personenbezogenen Daten, ggf. Absicht für eine Übermittlung an Drittland, Dauer der Speicherung, Recht auf Auskunft, Berichtigung, Löschung, Einschränkung, Widerruf und Beschwerde bei einer Aufsichtsbehörde.

- Ist die Bereitstellung der personenbezogenen Daten gesetzlich oder vertraglich vorgeschrieben?
- Ist Profiling vorhanden? Wenn ja, ist eine Benachrichtigung über die involvierte Logik und Tragweite erfolgt?
- Ist die Datenschutzerklärung leicht verständlich und jederzeit leicht auffindbar?

Recht auf Auskunft (7)

Vorgabe der DSGVO: Die betroffene Person hat gegenüber dem Verantwortlichen ein Recht auf Auskunft auf folgende Informationen: Verarbeitungszwecke, Kategorien der personenbezogenen Daten, Empfänger oder Kategorien von Empfängern (Drittländer, Organisationen), geplante Speicherdauer, Recht auf Berichtigung und Löschung, Beschwerderecht, Herkunft der Daten, wenn die personenbezogenen Daten nicht bei der betroffenen Person erhoben wurden, automatisierte Entscheidungsfindung einschließlich Profiling Artikel 15, Abs.1 geeignete Garantien bei Datenübertragung an ein Drittland Artikel 15, Abs.2 sowie Kopie der personenbezogenen Daten. Artikel 15, Abs. 3.

Resultierende Herausforderung: Betroffene Personen können von einer Auskunftsanfrage Gebrauch machen. Der Verantwortliche muss diese Anfrage schriftlich oder elektronisch beantworten können. Außerdem muss der Verantwortliche alle vertretbaren Mittel nutzen, um die Identität einer Auskunft suchenden betroffenen Person zu überprüfen. Bei begründeten Zweifeln an der Identität kann der Verantwortliche zusätzliche Informationen anfordern. Bei mangelnder Identifizierbarkeit der betroffenen Person kann der Verantwortliche die Auskunft verweigern.

Technischer Lösungsansatz: Um diese Herausforderung technisch zu unterstützen, sind flexible Schnittstellen notwendig, die es ermöglichen bestehende Daten aus dem System abzufragen. Am besten eignen sich bekannte Standardschnittstellen, wie REST, um Daten explizit aus dem System abzufragen. Demzufolge sind Standardabfragen festzulegen, die relevante Informationen (siehe Vorgabe der DSGVO) extrahieren. Die Informationen sind dann mittels Text und ggf. Bilder an den Nutzer auszuweisen. Möchte der Nutzer lediglich bestimmte Auskunftsinformationen erlangen, sind Auswahlfunktionen bereitzustellen. Abhängig von der Auswahl werden dann nur entsprechende Informationen geliefert.

Checkliste:

- Bietet das System eine Auskunftsmöglichkeit über die betroffene Person und die mit ihr verbundenen Daten an?

- Bietet das System eine Identifizierung der Auskunft suchenden betroffenen Person?

Recht auf Löschung (9)

Vorgabe der DSGVO: Ein Nutzer kann die Löschung von personenbezogenen Daten verlangen, sofern sie ihn betreffen. Die Verantwortlichen sind verpflichtet dem nachzukommen, sobald einer der Gründe gemäß Artikel 15 Abs. 1 a-f genannten Gründe vorliegt. Auch der Widerruf einer Einwilligung zählt hierzu. Weiterhin besagt der Artikel 15 Abs. 2, dass das Verlangen nach Löschung nach Möglichkeit an weitere Verantwortliche weiterzuleiten ist. Zudem existieren Ausnahmen, unter denen der Artikel nicht gilt Artikel 15 Abs. 3.

Resultierende Herausforderung: Die DSGVO verlangt eine Funktion zur Löschung von personenbezogenen Daten. Demnach müssen betroffene jederzeit die Möglichkeit besitzen eine Löschung ihrer Daten anordnen zu können. Außerdem ist sicherzustellen, dass eine Weiterleitung der Löschung an weitere Verantwortliche möglich ist.

Technischer Lösungsansatz: Beim Recht auf Löschung ist eine Schnittstelle bereitzustellen, die eine nachträgliche Löschung von personenbezogenen Daten ermöglicht. Daten von einzelnen müssen abfragbar und separat löschbar sein. Eine nachträgliche Reproduzierbarkeit der Daten ist nach der Löschung nicht zulässig.

Checkliste:
- Ermöglicht das System die Löschung von Benutzerdaten und Accounts?
- Können die Daten nach der Löschung wiederhergestellt werden?

3.3 Pflichten des Verantwortlichen (III)

In einem dritten Teilbereich sieht die DSGVO verschiedene Pflichten vor, denen eine verantwortliche Stelle nachkommen muss. Hierzu gehören die *Mitteilungspflicht (12)* und die auf Datenschutz optimierte Voreinstellung *Privacy by Default (13)*. Wir werden in diesem Abschnitt nur das zweite Muster skizzieren.

Privacy by Default (13)

Vorgabe der DSGVO: Durch geeignete technische und organisatorische Maßnahmen soll sichergestellt werden, dass die Voreinstellungen eines Dienstes die Benutzer bei der Erhebung, Verarbeitung, Speicherung und Weitergabe von personenbezogenen Daten nicht bevormundet. Dies wird häufig als Privacy by Default bezeichnet. Artikel 25 Abs. 2.

Resultierende Herausforderung: Die Erhebung, Verarbeitung, Speicherung und Weitergabe muss technisch auf jede relevante Nutzersituation einstellbar sein. Erst dadurch sind je nach Nutzersituation variable datenschutzfreundliche Voreinstellungen möglich.

Technischer Lösungsansatz: Im Kapitel Allgemeine Muster wurde durchgehend die Empfehlung gegeben, Daten mit zusätzlichen Attributen zu versehen. Beispielsweise, beim Muster *Zweckbindung* empfehlen wir zu jedem personenbezogenen Datum ein zusätzliches Attribut Zweck zu speichern. Anhand dessen kann eine attributbasierte Zugriffskontrolle die Zweckbindung technisch garantieren. Beim Muster Speicherbegrenzung empfehlen wir ein Zeitattribut, das die Speicherdauer repräsentiert. Sind die Daten mit solchen Attributen versehen, dann ist es technisch möglich auch datenschutzfreundliche Ausprägungen über diesen Attributen zu definieren. So können Daten standardmäßig zunächst mit dem generischen Attribut Datenablage versehen werden, sodass kein Verarbeitungsprozess auf diese zugreifen kann, da diese erstmal nur zu Ablage gedacht sind. Ähnliches ist auch über das Zeitattribut möglich, das je nach Art des Datums eine Lebensdauer bestimmt, ist diese abgelaufen, kann eine weitere Bearbeitung nicht mehr erfolgen. Das Attribut sollte individuell von der betroffenen Person eingestellt werden können. Die beiden Attribute sind exemplarisch zu verstehen. Der Betreiber eines Dienstes muss die Anforderungen dieses Musters bereits während des Entwurfs des Datenmodells berücksichtigen und passende Attribute mitsamt deren Ausprägungen definieren.

Checkliste:

- Verfügt das System über geeignete Steuerungsattribute, die die Daten kennzeichnen?
- Können Nutzer Einstellungen zur Verarbeitung von personenbezogenen Daten flexibel vornehmen?
- Werden Nutzer durch bestimmte Systemeinstellungen nicht bevormundet?

3.4 Bewertung der Muster

Abgeleitet von der vorhergegangenen Analyse können Abhängigkeiten zwischen den 13 Mustern identifiziert werden (siehe Tab. 1). Das Muster *Transparenz und Nachvollziehbarkeit (1)* weist eine hohe Abhängigkeit auf. Das Muster verfügt über Beziehungen zu den Mustern *Zweckbindung (2), Datenminimierung (3), Richtigkeit (4), Informationspflicht (6), Recht auf Auskunft (7), Recht auf Datenübertragbarkeit (11) und Mitteilungspflicht (12)*. Analog dazu können weitere Abhängigkeiten festgestellt werden. Da diese Muster aus der DSGVO abgeleitet wurden, kann im Umkehrschluss eine hohe Abhängigkeit der einzelnen DSGVO-Artikel untereinander ermittelt werden. Dies bedeutet, dass auch die gemeinschaftliche Umsetzung der einzelnen technischen Maßnahmen zu empfehlen ist.

Schwerpunkt		I					II						III	
	Muster	1	2	3	4	5	6	7	8	9	10	11	12	13
I	1	-	x	x	x		x	x				x	x	
	2	x	-			x	x						x	x
	3	x		-		x								x
	4	x			-				x	x				
	5		x	x		-								x
II	6	x	x				-	x				x	x	
	7	x					x	-	x	x		x	x	
	8				x			x	-	x	x		x	
	9				x			x	x	-	x		x	
	10								x	x	-			
	11	x					x	x				-	x	
III	12	x	x				x	x	x	x		x	-	
	13		x	x		x								-
I = Allgemeine Muster, II = Rechte der betroffenen Person, III = Pflichten der Verantwortlichen														

Tab. 1: Abhängigkeiten zwischen den einzelnen Mustern

Durch die beschriebenen, technischen Lösungsansätze wird eine Teillösung zu Q2 angeboten. Zur Umsetzung der technischen Lösungsansätze, bietet die Checkliste eine problembezogene und zielorientierte Hilfestellung. Mit diesem Wissen kann die Q2 Fragestellung vollständig beantwortet werden.

4 Anwendungsbeispiel anhand des Forschungsprojekts EDV

Die Anwendbarkeit ausgewählter Muster wurde im datenschutzsensiblen Forschungsprojekt Einfaches Digitales Vergessen (EDV) evaluiert. Das EDV Projekt stellt einen Lösungsansatz bereit, der die Grundsätze der DSGVO sowie die Rechten und Pflichten der betroffenen Personen und Verantwortlichen berücksichtigt. Das EDV-System ermöglicht das Austauschen von Dokumenten mittels einer App, wobei gezielt Zugriffsrechte und Zugriffsfristen eingestellt werden können und weder Betreiber des Dienstes

noch der Empfänger der Daten (sofern er sich an das vereinbarte Protokoll hält) die Daten unberechtigt lesen, länger aufbewahren oder Weiterleiten kann. Nachfolgend wird zu den zuvor erstellten Mustern, der EDV Lösungsansatz präsentiert.

Transparenz und Nachvollziehbarkeit (1): Vor der Benutzung des EDV-Systems wird eine Datenschutzerklärung angezeigt. Erfasste personenbezogenen Daten werden aufgelistet und an den Nutzer ausgewiesen. Außerdem werden die verwendeten Technologien sowie die Art und Weise der Verarbeitung beschrieben.

Zweckbindung (2): Das EDV-System gibt Verarbeitungszwecke in der Datenschutzerklärung an. Zu den erfassten Daten werden zusätzlich Verarbeitungszwecke gespeichert. Somit kann die Zweckbindung rückwirkend nachgewiesen werden.

Datenminimierung (3): EDV speichert lediglich Daten ab, die für den Austausch der Dokumente notwendig sind. Das System erlaubt eine Bearbeitung der Attribute. Durch die Nutzung von Docker-Containern können neue Strukturen umgehend eingespielt werden.

Speicherbegrenzung (5): Personenbezogene Daten werden mit einer Bearbeitungs- und Lesefrist im System abgelegt. Mit Ablauf der Frist werden die Informationen und ausgetauschten Dokumente gelöscht und sind somit nicht mehr zugänglich.

Informationspflicht (6): Vor Benutzung der EDV Anwendung wird der Nutzer informiert, wer die Daten verarbeitet und welche Daten verarbeitet und abgespeichert werden. Außerdem wird die betroffene Person bezüglich seiner Rechte informiert.

Recht auf Auskunft (7): Der Nutzer hat die Möglichkeit eine Auskunft seiner Daten zu verlangen. Auf Anfrage erhält der Nutzer alle relevanten Informationen, die ihn betreffen. Der Nutzer kann in der Anwendung einsehen, welche Daten im System abgelegt sind.

Recht auf Löschung (9): Mittels EDV-Systems hat der Nutzer die Kontrolle über seine Dokumente. Demnach kann er bisherige Informationen berichtigen und hochgeladene Dokumente komplett auch vor dem Ende der eingestellten Frist aus dem System entfernen. Dadurch ist auch kein Zugriff der Empfänger auf die Dokumente mehr möglich.

Privacy by Default (13): Das EDV-System bietet von Grund auf benutzerfreundliche Datenschutzeinstellungen. So ist beispielsweise das Benutzerprofil zunächst als privat angelegt und öffentlich nicht sichtbar. Der Nutzer kann in den Einstellungen frei entscheiden ob das Profil auch öffentlich zugänglich sein soll.

5 Zusammenfassung und Ausblick

Die Forschungsfrage Q1 konnten wir durch die Aufstellung von 13 Mustern positiv beantworten. Es ist also möglich problemorientiert und musterbasiert Lösungsansätze zu

technischen Anforderungen, die im Rahmen der DSGVO entstehen, zu erstellen. Diese Ansätze sind auf verschiedene Systeme übertragbar, wir konnten dies exemplarisch durch die Anwendung im EDV Projekt zeigen. Durch den klaren Aufbau der Muster können diese als Leitfaden oder Nachschlagewerk genutzt werden. Die Muster ermöglichen eine anwendungsfallspezifische Entwicklung von Lösungsansätzen auf Basis der DSGVO. Dabei erlauben es die Muster Herausforderungen und technische Lösungsansätze zu erkennen (Q2). Dies wird weiterhin unterstützt, durch den strukturierten Aufbau der Muster und die Einteilung in drei Kategorien: 1) Muster, die sich allgemein aus der DSGVO ableiten lassen; 2) Muster, die den Rechten der betroffenen Person zugeordnet werden können und 3) Muster, die den Pflichten des Verantwortlichen geschuldet sind. Zur vereinfachten Umsetzung dienen die den Mustern zugeordneten Checklisten. Verantwortliche werden dadurch besser in die Lage versetzt, technische Dienste konform zur Gesetzgebung umzusetzen.

Wie bereits zu Beginn des Artikels erwähnt wird die Digitalisierung weiter voranschreiten. Daher wird der Bereich Datenschutz und Datensicherheit eine immer wichtigere Rolle in einnehmen. Wir sehen daher den Bedarf weitere Anforderungen aus der DSGVO explizit zu beschreiben und zusätzliche, musterbasierte Lösungsansätze zu entwickeln. Auch wenn die beschriebenen Checklisten und die Strukturierung bereits in diese Richtung gehen, planen wir bezugnehmend zu Q2 eine vereinfachte Anwendung der Muster zu unterstützen. Dazu planen wir die Entwicklung eines web-basierten, interaktiven Musterkatalogs. Hierzu wollen wir künftig auch einen Fragekatalog entwickeln, der es Unternehmen ermöglicht, DSGVO-basierte Anforderungen automatisch anhand ihrer Anwendungsfälle (Systeme) zu erkennen und Lösungsansätze nach Bedarf abzuleiten. Ein weiteres zukünftiges Forschungsfeld sehen wir in der Identifikation von allgemeinen Datenschutzmustern, die Lösungsstrategien unabhängig von einzelnen Gesetzesvorgaben beschreiben. Hieraus könnte künftig auch ein erweiterter Musterkatalog entstehen, der hierarchisch aufgebaut ist und indem auch nach bestimmten Gesetzen gefiltert werden kann (beispielsweise nach der DSGVO). Neben einer gesetzbasierten Erweiterung des Musterkatalogs erscheint das Themenfeld Privacy by Default besonders zur weiteren Untersuchung geeignet. Als einen der nächsten Schritte planen wir daher diesen Bereich genauer zu untersuchen, um weitere Datenschutzmuster zu identifizieren.

Diese Arbeit entstand im Forschungsprojekt EDV (Förderkennzeichen 01MT17009A), gefördert durch das BMWi in der Förderlinie SmartData.

Literaturverzeichnis

[Ah07] Ahern, S., et. al.: Over-Exposed? Privacy Patterns and Considerations in Online and Mobile Photo Sharing. ACM, New York, S. 357-367, 2007.

[Al95] Alexander, C. et. al.: Eine Muster-Sprache. Städte, Gebäude, Konstruktion. Löcker Verlag, Wien, 1995.

[APOR18] Alpers, S.; Pilipchuk, R.; Oberweis, A.; Reussner, R.: Identifying Needs for a Holistic Modelling Approach to Privacy Aspects in Enterprise Software Systems. ICISSP: S. 74-82, 2018.

[Be02] Berry, C.: 2002. J2EE Design Patterns Applied. Real World Development with Pattern Frameworks, Wrox, Birmingham, 2002.

[BH04] Blakley, B.; Heath, C.: Security Design Patterns. The Open Group, Vereinigtes Königreich, 2004.

[Bu04] Buschmann, F. et. al.: Pattern-Oriented Software Architecture. A System of Patterns, Wiley, Chichester, 2004.

[BMR18] Beyerer, J., Müller-Quade, J. & Reussner, R.: Karlsruher Thesen zur Digitalen Souveränität Europas. Datenschutz Datensicherheit 42:5: S. 277-280, 2018.

[Dw08] Dwork, C.: Differential Privacy: A Survey of Results, Springer, China, 2008.

[Fo03] Fowler, M.: Patterns of Enterprise Application Architecture. Addison-Wesley, Boston, 2003.

[Ga08] Gamma, E. et. al.: Entwurfsmuster. Elemente wiederverwendbarer objektorientierter Software, Addison-Wesley, München, 2008.

[He04] Hevner, A. et. al.: Design Science in Information Systems Research. MIS Quarterly 28:1, S. 75-105, 2004.

[HW04] Hohpe, G.; Woolf, B.: Enterprise Integration Patterns. Designing, Building, and Deploying Messaging Solutions, Addison-Wesley, Boston, 2004.

[Hu17] Huth, D.: A Pattern Catalog for GDPR Compliant Data Protection. In: Proc. 10th Int. Conf. on IFIP of Enterprise Modelling. Leuven, S. 34-40, 2017.

[Ro06] Romanosky, S. et. al.: Privacy Patterns for Online Interactions. In.: Proc. of the 2006 conference on Pattern languages of programs. ACM Press, New York, S.1-15.

[Sc03] Schumacher, M.: Security Patterns and Security Standards. With Selected Security Patterns for Anonymity and Privacy, TU Darmstadt, Darmstadt, 2003.

[Sc04] Schümmer, T.: The Public Privacy. Patterns for Filtering Personal Information in Collaborative Systems, Universität Hagen, Hagen, 2004.

[Sc18] Schild H.: BeckOK Datenschutzrecht, C.H. Beck, München, 2019.

[SNL05] Steel, C.; Nagappan, R.; Lai, R.: Core Security Patterns: Best Practices and Strategies for J2EE, Web Services, and Identity Management, Prentice Hall, Saddle River, 2005.

[Sw02] Sweeney, L.: k-Annonymity: A Model for Protecting Privacy, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2002.

[Wi12] Wilder, B.: Cloud Architecture Patterns. Develop Cloud-Native Applications, O'Reilly, Peking, 2012.

The Layered Privacy Language Art. 12 - 14 GDPR Extension – Privacy Enhancing User Interfaces

Armin Gerl[1], Bianca Meier[2]

Abstract: On 25th May 2018, the EU-wide General Data Protection Regulation (GDPR) came into force in order to strengthen the rights of Data Subjects. Although the GDPR specifies the required information, which has to be presented to a Data Subject, it can still be argued for a lack of transparency due to unfavorable presentation of the privacy policy. Furthermore, no systematic approach for the enforcement of privacy policies in technical systems is deployed. These issues are tackled by the both human- and machine-readable Layered Privacy Language (LPL), which models legal privacy policies. This work introduces an extension for LPL to comply with Art. 12 - 14 GDPR. Additionally, user interface prototypes will be introduced to allow the creation of LPL privacy policies by the Data Protection Officer as well as a structured presentation of the LPL privacy policy for web-applications.

Keywords:

GDPR; Informed Consent; Layered Privacy Language; Privacy Policy; Privacy Management

1 Introduction

Where personal data is collected or processed users (Data Subjects) have to be informed about it by the privacy policy. Although this document is essential and contains all legally required information, users often do not read the privacy policies [Bi15] [St16]. This behaviour has various reasons, for example complexity, legal language or the length of the privacy policy [An11a]. Thus, the presentation of the privacy policy has to be reconsidered. McDonald and Cranor analyzed the cost of reading privacy policies in their study. The result was the following: If every American internet user reads all privacy policies, which are displayed to him, the whole nation needs for reading 54 billion hours per year. Breaking down this sum, every American citizen would require 40 minutes every day reading privacy polices [MC08]. As a result of this great expenditure of time, many users agree/consent to privacy policies without any understanding. The GDPR, which intends to strengthen the rights of Data Subjects, e.g. by requiring free and informed consent [GD16, Art. 7], seems not to have any noticeable effect on this behaviour. To avoid unknown processing of personal data, the Data Subject has to understand the contents of the privacy policy, which is non-trivial. Due to the complexity of the GDPR and its definition of information that has to be provided to the Data Subject [GD16, Art. 12 - 14], also the creation of GDPR

[1] University of Passau, Chair of Distributed Information Systems, Passau, Germany armin.gerl@uni-passau.de
[2] University of Passau, Chair of Distributed Information Systems, Passau, Germany bianca.meier@uni-passau.de

doi:10.18420/inf2019_51

compliant privacy policies is challenging for the Controller, which can be supported by the Data Protection Officer (DPO). Because, no uniform approach of creating and handling is available, the management of privacy policies is a tedious and time-consuming task which may be individual for each DPO and Controller. Furthermore, this results in various structures, wordings and presentations of privacy policies hindering the understanding for the Data Subjects.

To tackle this challenge, we propose a systematic computer science approach to create and present privacy policies in a unified way utilizing the Layered Privacy Language (LPL). Based on LPL an overarching framework enables the privacy-preserving processing of personal data directly based on the decisions of the users. Therefore, users negotiate and agree/consent to a LPL privacy policy, which represents the individuals' privacy settings. This personalized privacy policy is considered for each processing, i.e. processing of personal data is restricted to specific purposes. Thus, the users' decisions on its privacy are directly influencing if and how its personal data is processed [Ge18b]. Furthermore, the human- and machine-readable LPL enables a systematic creation of privacy policies, which can be verified for completeness, and the structured presentation of privacy policies. To comply with the requirements for the contents of a privacy policy, LPL is extended according to Art. 12 - 14 GDPR [GP18a]. We detail how this LPL extension complies to the GDPR, such that legal privacy policies can be modelled.

The main contribution of this work consists of the introduction of the *LPL Policy Creator*, which intends to support the Controller with the creation and management of privacy policies. And the extension of the *LPL Policy Viewer*, presenting users the required information, based upon previous work [GP18b] [Ge18a] to comply with the GDPR requirements for privacy policies. As a result users can perceive standardized policies including all necessary information in a layered approach [Gr18]. The remaining of the paper is structured as follows. Section 2 reviews related work regarding other privacy languages and the visualizations of privacy policies. The extension of LPL to the Art. 12 - 14 of the GDPR is detailed in section 3. The *LPL Policy Creator* is introduced in section 4. Section 5 presents the updated *LPL Policy Viewer* for the presentation of the privacy policy to the user. Lastly, section 6 concludes this work and gives an outlook.

2 Related Work

Next to LPL other privacy languages have been proposed to enhance the privacy experience, which we will shortly describe and compare to, to show the strengths of LPL.

The *Privacy Preferences Project*, short P3P, is standardized by the the World Wide Web Consortium (W3C) [CAG02]. P3P intended to provide privacy policies in a standardized format and therefore enable automatic processing of them. Naturally, P3P does not consider GDPR, because it has been proposed before GDPR. P3P models privacy policies in XML, which then is provided by the website to the user via the browser. To model privacy policies

a pre-defined vocabulary is used, which only allows for a restricted extent the modelling of real privacy policies, e.g. the vocabulary for purpose is fixed. In contrast LPL does not use a pre-defined vocabulary for its elements. To support the user with the decision if the provided P3P privacy policy complies with its personal privacy preferences the *A P3P Preference Exchange Language (APPEL)* is introduced [CAG02]. Processing both the personal privacy preferences of the user and the P3P privacy policy provided by the website, the browser plugin *Privacy Bird* [CGA06] visualizes the fulfillment of the users privacy preferences via an icon. Three different icons exist: A green bird, which tells the user, that his personal privacy settings are consistent with the privacy policy. A red bird indicates that they are not consistent. Lastly, a yellow bird indicates that the tool is unable to retrieve a privacy policy from the website. Thus, only a few indications are given to user, but no further information, interaction, or possibility for consent management are given.

The privacy language PrimeLife (PPL) [An11b] intends to handle access control and data usage at the same time. The newest guidelines of the GDPR will not be considered, because PPL was implemented before it. PPL comes with a user interface for the presentation and negotiation of privacy policies. To tackle the challenge that it is hard for the user to define its own privacy preferences, pre-configured levels of privacy are provided – 'Nearly Anonymous', 'Minimal Data' and 'Requested Data' – which can be chosen and changed by the user at any time [An11b]. To visualize the data processing, a dialog called *Send Data?* is proposed [An11b], which presents the user in a tabular visualization the collected data for each purpose. Additionally, data recipients are listed. The user interfaces enables the comparison to the users personal settings, but does not allow the negotiation of the content of the privacy policies. This is in contrast to LPL, which supports negotiation.

The *SPECIAL Project*, which was funded by European Union's Horizon 2020 research, proposes a GDPR compliant privacy dashboard [PRK17] based upon SPECIAL's Usage Policy Language [PB17]. The privacy dashboard provides a time-line consisting of each processing of data. Data items are divided in four groups: 'Data I provide', 'Data of me provided by others', 'Data of my behavior' and 'Inferred data about me'. This subdivision improves the transparency of data processing for the Data Subject. In addition to this time-line, the user interface informs about the privacy policy in a written way and third parties. It is important to notice, that the user can give/withdraw his consent to the processing for any purpose, which represents the negotiation of a privacy policy in LPL. Hereby, it is differentiated between required and non-required purposes. Required purposes have to be agreed upon by the user and are usually necessary for the service, thus they cannot be withdrawn from the privacy policy. On the other hand non-required purposes have to be consented to by the user.

3 LPL with Art. 12 - 14 Extension

LPL is intended to represent all privacy policy concerning processes including creation, negotiation, managing and enforcing of privacy policies [Ge18b]. Thus, LPL has to be

presentable in a human-readable way that supports free and informed consent, while personalization of the privacy policy is encouraged. Furthermore, LPL enables the enforcement of the privacy policy due to policy-based access control and de-identification mechanisms. Therefore, data can only be processed by authenticated and authorized entities in a, if necessary, de-identified way. To achieve this personal anonymization, pseudonymization methods, and privacy models are integrated.

This work focuses on the creation and presentation of privacy policies in the context of Art. 12 - 14 GDPR, such that LPL policies can be used within the European legal framework. For the presentation of LPL privacy icon capabilities and human-readable headers and descriptions with internationalization support have been introduced [Ge18a] [GP18b]. Therefore, Art. 12 - 14 GDPR has been analyzed and requirements have been derived. Comparing the original version of LPL [Ge18b] against those requirements it was found that the basic policy structure is full-filled, but several informative requirements are missing for which an extension has been proposed [GP18a]. Furthermore, LPL has been extended by pseudonymization capabilities, which are necessary in health care scenarii [GB19]. Within this work we consider LPL with all mentioned extensions (see Fig. 1) to cover its full extent for creation and presentation. Therefore, we reconsider the requirements defined by Gerl and Pohl [GP18a] and compare them to the updated LPL in the following.

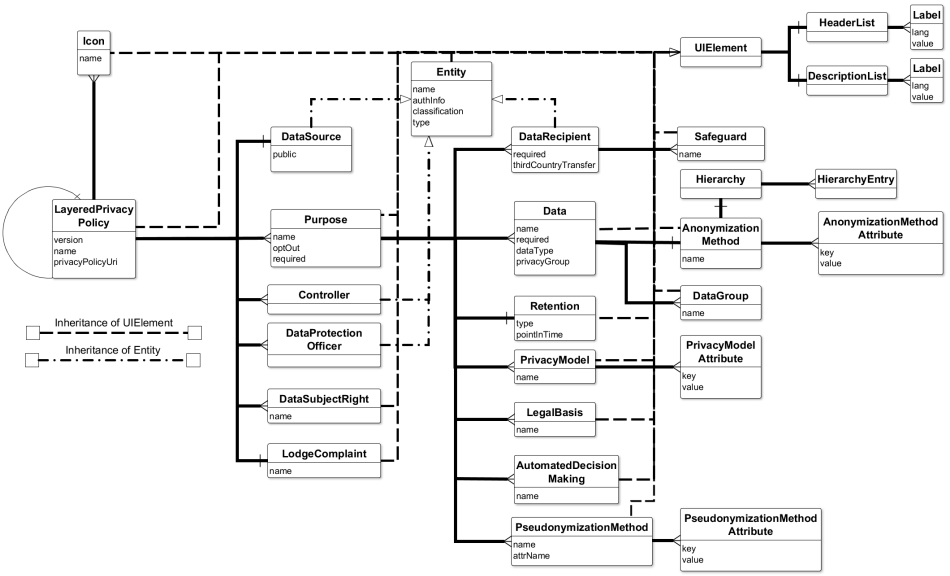


Fig. 1: The structure of the Layered Privacy Language [Ge18b] with the *User Interface Extension* [Ge18a], *Pseudonymization Extension* [GB19], and *Art. 12 - 14 GDPR Extension* [GP18a].

3.1 Comparison of LPL to Art. 12 - 14 GDPR Requirements

The main articles of the GDPR dealing with the requirements for privacy policies are Art. 12 - 14, which will be compared to the capabilities of LPL in the following (see Tab. 1).

In Art. 12 GDPR general provisions for the communication to the Data Subject, especially regarding transparency, are stated [GD16, Art. 12]. First of all it states that a privacy policy has to be provided in a clear and plain language [GD16, Art.12 (1) Sentence 1], which is enabled through the UIElement providing all key elements with human-readable headers and descriptions, which has been introduced in the *User Interface Extension* [Ge18a]. Furthermore, the privacy policy can be provided in a written or electronic form [GD16, Art.12 (1) Sentence 2], under which LPL falls as an electronic format. What cannot be covered by the LPL model itself is the realization of the Data Subject Rights [GD16, Art. 12 (2)], their response time [GD16, Art. 12 (3)], or the protection of the Controller from excessive Data Subject Rights requests [GD16, Art. 12 (5)], because this is concerning an overarching privacy framework using LPL. Realization of the semi-automatization of Data Subject Rights is hereby subject to future work. The last requirement derived of Art. 12 allows the usage of standardized icons [GD16, Art. 12 (7)], which are covered by LPL through the introduction of the Icon-element for privacy icons [Ge18a].

The following Art. 13 and Art. 14 have very similar content and will therefore be combined compared. Art. 13 describes the information that has to be provided where personal data are collected from the Data Subject [GD16, Art. 13] and Art. 14 describes the information that has to be provided before data is collected from the Data Subject [GD16, Art. 14]. Both articles demand that the identity of the Controller and its contact details are provided [GD16, Art. 13 (1)(a), Art. 14 (1)(a)], which is modelled in LPL as a set of Controller-elements to also consider Joint Controllers [GD16, Art. 26]. Furthermore, the contact details of the responsible DPO has to be provided [GD16, Art. 13 (1)(b), Art. 14 (1)(b)], which is covered by the DataProtectionOfficer-element of LPL. The purposes of the processing of personal data and their legal basis, including the legitimate interests [GD16, Art. 13 (1)(d), Art. 14 (2)(a)], have to be stated [GD16, Art. 13 (1)(c), Art. 14 (1)(c)], which LPL models a set of Purpose-elements each having a set of LegalBasis-elements. The Data Subject has to be informed about the collected data categories [GD16, Art. 14 (1)(d)] modelled by the DataGroup-element. The required personal data has to be communicated to the Data Subject [GD16, Art. 13(2)(e)], which is modelled by the Data-element having the required attribute. The data recipients for the personal data [GD16, Art. 13 (1)(e), Art. 14 (1)(e)], and if the data is transferred in a third country and the applied safeguards [GD16, Art. 13 (1)(f), Art. 14 (1)(f)] have to be provided, which is modelled by LPL as a set of DataRecipient-elements which have an attribute indicating a third country transfer and a set of Safeguard-elements if necessary. The storage period for the personal data has to be provided to the Data Subject [GD16, Art. 13(2)(a), Art. 14(2)(a)], which is modelled by the Retention-element of LPL. Furthermore, the Data Subject has to be informed about its Data Subject Rights [GD16, Art. 13(2)(b), Art. 14(2)(c)] and how to lodge a complaint [GD16, Art. 13(2)(d), Art. 14(2)(e)],

which is implemented by the DataSubjectRights-element and LodgeComplaint-element globally for a policy. The Data Subject has to be informed if automated decision-making is performed based on the personal data [GD16, Art. 13(2)(f), Art. 14(2)(g)], which is modelled for each purpose by the AutomatedDecisionMaking-element. The Data Subject has further to be informed about the possibility to withdraw consent [GD16, Art. 13(2)(c). Art. 14(2)(d)], this is implicitly modelled in LPL by the required attribute for the Purpose-element that allows the user to (withdraw) consent to a purpose. This concept is further extended to the Data-element and DataRecipient-element allowing for several personalization options. Lastly, the Data Subject has to be informed about the source of personal data and if this source is publicly available [GD16, Art. 14(2)(f)], which is modelled b the DataSource-element with the public attribute denoting a public source.

Thus, LPL shows the capabilities to model all required information required by Art. 12 - 14 GDPR, whereas the implementation of the Data Subject Rights and their execution have to be considered by the overarching privacy framework utilizing LPL. It may be argued that the information stating that withdrawing consent is possible, may be modelled explicitly, but in LPL we consider personalization of the privacy policy as a feature that should cover this requirement and allow the user to influence its personal data 'from consent to processing'.

GDPR		**LPL**
Article	**Requirement**	**Implementation**
Art. 12(1) Sentence 1	Clear and Plain Language	UIElement
Art. 12(1) Sentence 2	Written or Electronic Information	LayeredPrivacyPolicy
Art. 12(2)	Data Subject Rights Realization	Framework
Art. 12(3)	Response Time	Framework
Art. 12(5)	Excessive Requests	Framework
Art. 12(7)	Standardized Icons	Icon
Art. 13(1)(a), Art. 14(1)(a)	Contact Details of Controller	Controller
Art. 13(1)(b), Art. 14(1)(b)	Contact Details of DPO	DataProtectionOfficer
Art. 13(1)(c), Art. 14(1)(c)	Purpose and Legal Basis	Purpose; LegalBasis
Art. 13(1)(d), Art. 14(2)(b)	Legitimate Interest	LegalBasis
Art. 14(1)(d)	Categories of Personal Data	DataGroup
Art. 13(1)(e), Art. 14(1)(e)	Recipients of Personal Data	DataRecipient
Art. 13(1)(f), Art. 14(1)(f)	Third Country Transfer	DataRecipient; Safeguard
Art. 13(2)(a), Art. 14(2)(a)	Storage Period	Retention
Art. 13(2)(b), Art. 14(2)(c)	Information: Data Subject Rights	DataSubjectRights
Art. 13(2)(c). Art. 14(2)(d)	Information: Withdraw Consent	Purpose
Art. 13(2)(d), Art. 14(2)(e)	Information: Lodge a Complaint	LodgeComplaint
Art. 13(2)(e)	Information: Required Data	Data.required
Art. 14(2)(f)	Source of Personal Data	DataSource
Art. 13(2)(f), Art. 14(2)(g)	Automated Decision-Making	AutomatedDecisionMaking

Tab. 1: Overview of the implementation of the legal requirements for privacy policies according to Art. 12 - 14 GDPR by the Layered Privacy Language.

4 LPL Policy Creator

One of the tasks of the Controller is to create and manage privacy policies. Hereby, it can be a challenge to keep track of the fulfillment of all requirements given by the GDPR. The *LPL Policy Creator* is a prototype implementation, which supports the creation process for LPL privacy policies, while support for compliance with GDPR is given. For a better understanding we assume the following scenario.

The company 'Shopping Worldwide' operates a web shop. To comply with GDPR the company has to create a privacy notice according to Art. 13 and Art. 14 GDPR which will be integrated in the privacy policy to inform the users about the processing of their personal data. The company uses the collected data for the non-required purpose 'Marketing' and the required purposes 'Billing' and 'Research'. The later one requires the data 'sex', 'age' and 'salary-class' and the non-required data elements 'education' and 'work-class' belong to the last one. The collected data is processed by the data recipient 'internal', which is the company itself. The other data recipient, 'external', denotes that the data is analyzed by a third party (which should be denoted in detail for a real-life policy). The data recipient 'internal' is required, because the company itself has to process the data to provide the web shop. In contrast, the data recipient 'external' is non-required, such that the user has to actively consent to it. In this scenario we denote a fictional legal basis 'National Research Initiative' as the legal basis of the processing. Also a fictional retention with the deletion type 'INDEFINITE' was created. No de-identification (anonymization, pseudonymization, or privacy model) will be defined for this purpose, also no automated decision-making will be conducted. Further description of other purposes is omitted for the scope of this paper. These requirements can be modelled with the *LPL Policy Creator* using an interactive user interface (see Fig. 2). Individual elements are detailed in the following.

Header The *Header* provides three different functions: Add a new layer to the current privacy policy, reset the whole created privacy policy and the possibility to import a LPL privacy policy. Creating a new LPL policy layer enables further detailing or restriction of the policy, e.g. the user consented policy defines that data can be used for marketing, then an additional internal privacy policy layer can be added to further specify that only specific data is accessible by specific roles or departments within the company. Therefore, a LayeredPrivacyPolicy-element includes a set of UnderlyingPrivacyPolicies-elements, which are LayeredPrivacyPolicy-elements.

Policy Header The *Policy Header* is separated into general settings and the *Privacy Icon Overview* [Ge18a]. The general settings, accessed with the button 'Edit', allow to alter the language for international support. Additionally, a link (URL) to the regular legal privacy policy can be set, to comply with common practices. Also other elements of the policy can be set within the header, e.g. information about the Data Subject Rights or that the Data Subject can lodge a complaint. The *Privacy Icon Overview* enables the addition of privacy

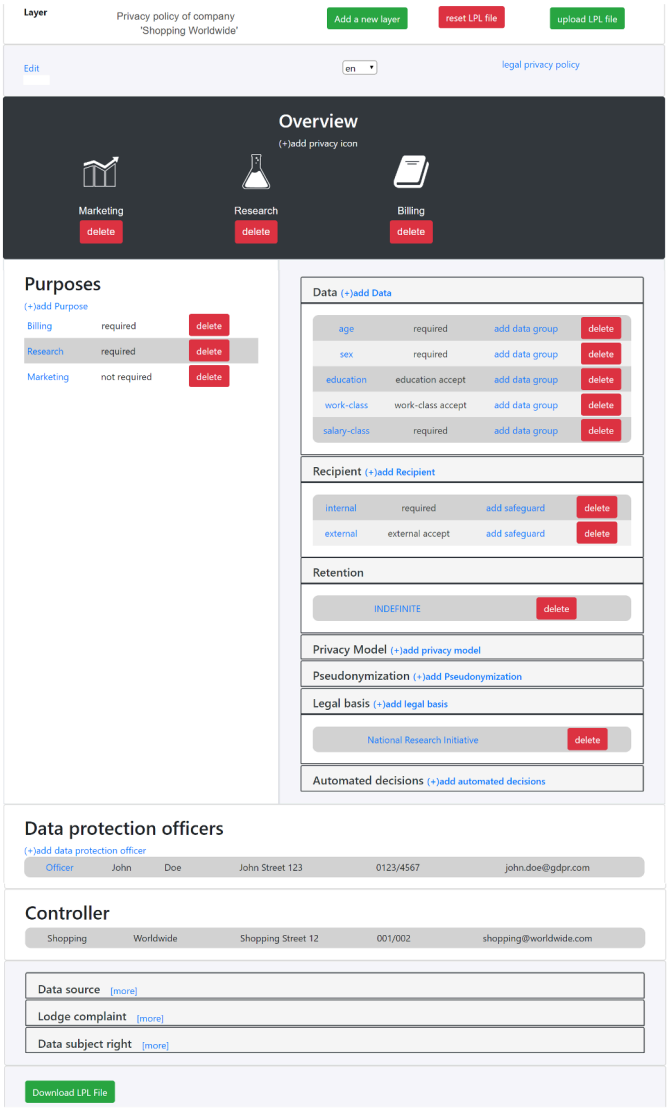


Fig. 2: *LPL Policy Creator* example creating a LPL privacy policy with the purposes 'Billing', 'Research', and 'Marketing'. The purpose 'Research' is selected detailing further information on, e.g. the processed data, the data recipient, or retention. Furthermore, information on the Data Protection Officer, Controller, as well as required information for the Data Subject is presented.

icons [GD16, Art. 12(7)] to the privacy policy. These icons are intended to support the understanding of privacy policies by providing a quick overview over the processing of personal data [Ge18a]. The Controller is intended to select from a specified list of icons. Because no official privacy icons have been implemented, we use self-defined privacy icons as placeholders, until a European standard is in place. Based on the given scenario, icons for the purposes 'Research' and 'Marketing' would be specified (see Fig. 2).

Purpose Overview The *Purpose Overview* allows the management (create, update, delete) of purposes for the current privacy policy. The created purposes are listed, showing if they are required or not. For example the purpose 'Research' is required and therefore a indicating text-field is shown. The non-required purpose 'Marketing' is similarly denoted (see Fig. 2). For each purpose additional settings can be conducted by selecting it, e.g. adding a descriptive text for the purpose.

Purpose Detail For each purpose various information can or has to be provided. Therefore, for every purpose a set of data, set of data recipients, set of legal basis and retention has to be provided. Furthermore, pseudonymization and privacy models may optionally be defined, as well as information about automated decision-makings. Furthermore, for each data element an anonymization method can be defined, to allow for fine-grained de-identification rules, e.g. a postal-code may be anonymized for a marketing purpose using suppression. Both for data and data recipients it can be defined if they are required or optional, such that the user can decide on what data is processed for which purpose by whom. The data recipient can hereby be a company, department, role, or individual, while data represents the actual attribute, e.g. column of a table in a relational database. If a data recipient is not covered by the GDPR, e.g. a company in the USA, then safeguards have to be implemented and specified for the data recipient. Retention of data can be set as a fixed date, in relation of the ending of the purpose, or indefinitely. For privacy models, which define privacy guarantees for the whole data-set and not only a single record, common privacy models are supported, e.g. *k-Anonymity* [SS98] or *Differential Privacy* [Dw06]. Because the selection of the appropriate privacy model is non-trivial, we intend to support decision of with a questionnaire-based wizard in future works. Similarly, pseudonymization method can be defined to tokenize personal information like the name, e.g. hashing [Aa13].

General Information In the *General Information* section of the *LPL Policy Creator* common information on the privacy policy has to be created. This includes the contact details of the DPO or several DPOs iff applicable, the responsible Controller or a set of Controllers to allow for Joint Controllers [GD16, Art. 26], information on the data source i.e. the Data Subject after the acceptance of the privacy policy, and information on how to lodge a complaint as well as Data Subject Rights.

Footer After finishing the creation of a LPL privacy policy, it can be stored as a JSON or XML file, which allows for the integration in services using LPL and being presented by the *LPL Policy Viewer*. Furthermore, this functionality ensures the re-usability of privacy policies, which were created with the vocabulary of LPL.

5 LPL Policy Viewer

Next to the *LPL Policy Creator*, the first iteration of the *LPL Policy Viewer* [GP18b] with its Privacy Icon Overview [Ge18a] has been extended to incorporate all elements of the *LPL Art. 12 -14 GDPR Extension* as well as fine-grained consent management [GMB19]. The *LPL Policy Viewer* is hereby intended to give the user a fast overview over the processing of its personal data, while all necessary information is provided due to *layering*. In order to make the information even more comprehensible, the Visual Information Seeking Approach (VISA) – Overview first, zoom and filter, details on demand – is applied [Sh96]. Furthermore, the user is enabled to personalize the privacy policy by consenting to non-required elements i.e. purpose, data and data recipient. Further support for influencing the anonymization settings is anticipated for future work.

The initial concept of the *LPL Policy Viewer*, which consists of an overview over the purposes of the processing of the personal data both using privacy icons and an purpose overview, has not been altered. Only after interacting with the *LPL Policy Viewer* additional information is revealed. This is implemented by the so-called 'information overload' for the user, therefore it is important for the user to prepare the information in such a way that he can quickly and easily find relevant details without explicitly searching for them [MMG02]. This is especially important, because privacy policies are in general complex, such that the abstraction without loss of required information is essential.

Compared to the *LPL Policy Creator* the *LPL Policy Viewer* is structured similarly, but it lacks the functionality to create new elements, e.g. purposes, for the policy. Instead, it first presents the user with an overview of the privacy policy, then allows for browsing for specific information, e.g. the data recipient for personal data of a specific purpose or information on how to lodge a complaint. Due to the extension of LPL all required information according to Art. 12 - 14 GDPR is provided. But it should be noted that for Art. 13(2)(c) and Art. 14(2)(d) GDPR, which specify that the Data Subject has to be informed about his right to withdraw consent if the legal basis of processing was consent, is implemented implicitly by the *LPL Policy Viewer*. The user is informed to be able to withdraw consent using check-boxes next to the purpose, data or data recipient element, which is only possible for elements that have the 'required' attribute set to 'false'. This enables the user to personalize its privacy policy at any time, but the personalization does not only affect the privacy policy but also the corresponding business processes due to the machine-readability of LPL. Thus, the withdrawal of consent to specific data fields removes them also from being processed for the specific purpose, which allows personalized applications. Furthermore, the user can

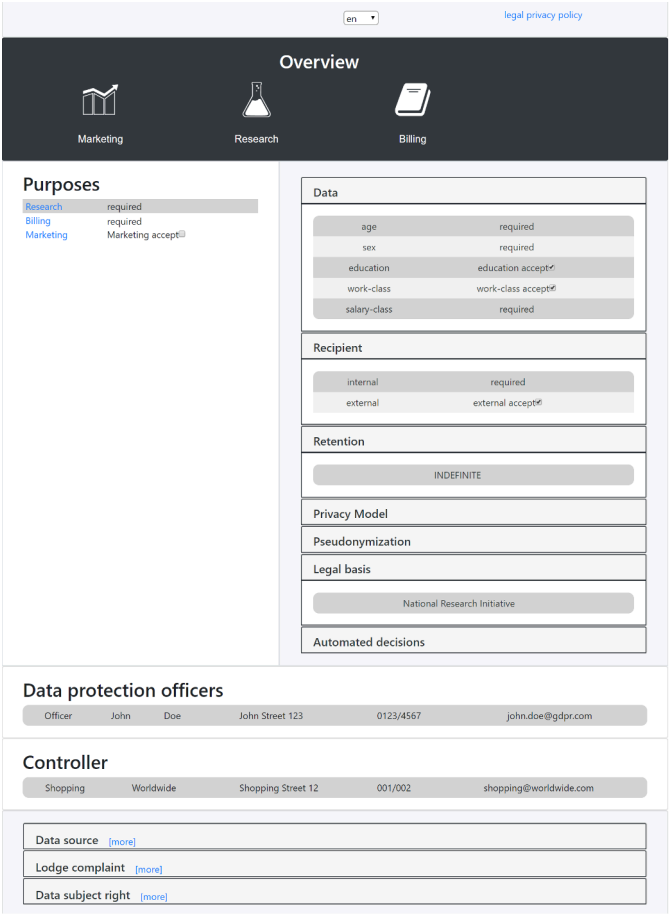


Fig. 3: Example for the LPL Policy Viewer

experience notable changes to the application [GMB19], which may give certainty that personal data is processed only according to its personal privacy policy.

In the following, we will describe the features provided by the *LPL Policy Viewer* based on the previously detailed example scenario (see Fig. 3). Due to the same base structure as the *LPL Policy Creator*, we will only highlight features that are essential for the presentation and negotiation of the privacy policy in the following.

Policy Header The *Policy Header* the user is presented the *Privacy Icon Overiew* [Ge18a], displaying individual icons for the purposes 'Marketing', 'Research', and 'Billing' giving the user an overview over the processing of its personal data 'at a glance'. Assuming

standardized privacy icons will be introduced, the GDPR stating that privacy icons may be provided in combination with the remaining privacy policy information, will be fulfilled [GD16, Recital 60]. Furthermore, the *Policy Header* provides the user with the capability to change the language of the privacy policy allowing for internationalization support. Lastly, the link for the legal privacy policy is provided to comply with the current state-of-the-art representation of privacy policies.

Purpose Overview On the left side of the user interface the *Purpose Overview* is located giving the user a textual list of the purposes for the processing of personal data. For each purpose it is indicated whether it is required and therefore necessary to agreed upon, or optionally accepted via interacting with the provided check-box. To comply with GDPR, the withdrawal of consent is as easy as giving it by clicking on the check-box [GD16, Art. 7(3) Sentence 3]. Clicking on the name of the purpose, e.g. 'Research', its details are given within the *Purpose Detail* section of the user interface. This interaction corresponds to the *Visual Information Seeking Approach* denoting that further details should be given on-demand after filtering [Sh96].

Purpose Detail Depending on the selected purpose additional information is shown within the *Purpose Detail*. Thus the user gets informed about which of its personal data is processed by which data recipients, when the data is deleted, if the data is protected by any de-identification techniques, and if any automated decision-makings are conducted for the specific purpose. Furthermore, the user can withdraw its consent to specific data or the processing by specific data recipients, iff they are not required. Given our example scenario, the data fields 'education' and 'work-class' are not required as well as the data recipient 'external', therefore consent can be withdrawn by interacting with the check-box (see Fig. 3).

General Information Lastly, in the *General Information* section of the user interface, the remaining required information regarding Art. 12 - 14 GDPR is represented, which has been missing in the first iteration of the user interface. The responsible DPO and Controller are hereby prominently presented, while information on Data Subject Rights or how to lodge are complained are accessible after interaction with the corresponding element. The reasoning behind this is, that the user should be aware of the Controller and the responsible Data Protection Office and their contact details before additional actions are taken, e.g. making use of a Data Subject Right.

6 Conclusion and Future Work

This work compared the current implementation of the Layered Privacy Language (LPL) with all its extension to the requirements given by in Art. 12 -14 GDPR for privacy policies

demonstrating full coverage. To demonstrate the coverage of the extension of LPL we introduced the *LPL Policy Creator* to support companies in the creation of privacy policies. Additionally, we extended *LPL Policy Viewer* to incorporate the extensions of LPL, allowing for a concise presentation of the privacy policy utilizing privacy icons and several interaction possibilities to enable fine-grained consent management.

Future works will extend the consent management pattern to incorporate the influence of anonymization properties for the Data Subject, as well as supporting with the selection of suitable de-identification methods. Also other user groups like children or elderly people have to be considered for future user interface concepts. Furthermore, the realization of Data Subject Rights as an semi-automated system utilizing LPL is subject of research, such that only minimal required actions from the DPO are necessary to respond. The goal is hereby to create a holistic approach to handle privacy intra and inter Controllers, while privacy guarantees can be given for Data Subjects.

Bibliography

[Aa13] Aamot, Harald; Kohl, Christian Dominik; Richter, Daniela; Knaup-Gregori, Petra: Pseudonymization of patient identifiers for translational research. BMC Medical Informatics and Decision Making, 13(1):75, Jul 2013.

[An11a] Angulo, J.; Fischer-Hübner, S.; Pulls, T.; Wästlund, E.: Towards usable privacy policy display & management-The primelife approach. In: Proceedings of the 5th International Symposium on Human Aspects of Information Security and Assurance, HAISA 2011. pp. 108–118, 2011.

[An11b] Angulo, Julio; Fischer-Hübner, Simone; Pulls, Tobias; König, Ulrich: HCI for Policy Display and Administration. In (Camenisch, Jan; Fischer-Hübner, Simone; Rannenberg, Kai, eds): Privacy and Identity Management for Life. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 261–277, 2011.

[Bi15] Bitkom: , Stimmen Sie den Aussagen voll / eher zu? Datenschutzerklärungen... https://de.statista.com/statistik/daten/studie/467075/umfrage/beurteilung-der-datenschutzerklaerungen-von-online-diensten-in-deutschland/, 2015.

[CAG02] Cranor, Lorrie Faith; Arjula, Manjula; Guduru, Praveen: Use of a P3P User Agent by Early Adopters. In: Proceedings of the 2002 ACM Workshop on Privacy in the Electronic Society. WPES '02, ACM, New York, NY, USA, pp. 1–10, 2002.

[CGA06] Cranor, Lorrie Faith; Guduru, Praveen; Arjula, Manjula: User Interfaces for Privacy Agents. ACM Trans. Comput.-Hum. Interact., 13(2):135–178, June 2006.

[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.

[GB19] Gerl, Armin; Bölz, Felix: Layered Privacy Language (LPL) Pseudonymization Extension for Health Care. In: Proceedings of MedInfo 2019. 2019.

[GD16] GDPR: , General Data Protection Regulation, April 2016. Regulation (EU) 2016 of the European Parliament and of the Council of on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

[Ge18a] Gerl, Armin: Extending Layered Privacy Language to Support Privacy Icons for a Personal Privacy Policy User Interface. In: Proceedings of Brithish HCI 2018. BCS Learning and Development Ltd., Belfast, UK, p. 5, 2018.

[Ge18b] Gerl, Armin; Bennani, Nadia; Kosch, Harald; Brunie, Lionel: LPL, Towards a GDPR-Compliant Privacy Language: Formal Definition and Usage. In (Hameurlain, Abdelkader; Wagner, Roland, eds): Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXVII. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 41–80, 2018.

[GMB19] Gerl, Armin; Meier, Bianca; Becher, Stefan: Let Users Control their Data – Privacy Policy-based User Interface Design. In: Human Interaction and Emerging Technologies 2019 - Proceedings of the 1st International Conference on Human Interaction and Emerging Technologies (IHIET 2019) conference. Université Côte d'Azur, Nice, France, August 2019.

[GP18a] Gerl, Armin; Pohl, Dirk: Critical Analysis of LPL according to Articles 12 - 14 of the GDPR. In: Proceedings of International Conference on Availability, Reliability and Security. ARES 2018, Hamburg, Germany, p. 9, August 2018.

[GP18b] Gerl, Armin; Prey, Florian: LPL Personal Privacy Policy User Interface: Design and Evaluation. In: Mensch und Computer 2018 - Tagungsband. Gesellschaft für Informatik e.V., Bonn, 2018.

[Gr18] Greger, Sebastian: , User-centred transparency design for privacy – Part I: The layered approach. https://sebastiangreger.net/2018/08/user-centred-transparency-design-the-layered-approach/, August 2018.

[MC08] McDonald, A. M.; Cranor, L. F.: The cost of reading privacy policies. I/S: A Journal of Law and Policy for the Information Society, 4, 2008.

[MMG02] Melgoza, Pauline; Mennel, Pamela A.; Gyeszly, Suzanne D.: Information overload. Collection Building, 21(1):32–43, 2002.

[PB17] P.A. Bonatti, S. Kirrane, I. Petrova L. Sauro E. Schlehahn: Deliverable D2.1 - Policy Language V1. Technical report, Scalable Policy-aware Linked Data Architecture For Privacy, Transparency and Compliance - SPECIAL, December 2017.

[PRK17] Philip Raschke, Axel Küpper, Olha Drozd; Kirrane, Sabrina: Designing a GDPR-compliant and Usable Privacy Dashboard. In: IFIP Advances in Information and Communication Technology. IFIP Summer School 2017, Springer, September 2017.

[Sh96] Shneiderman, B.: The eyes have it: a task by data type taxonomy for information visualizations. In: Proceedings 1996 IEEE Symposium on Visual Languages. pp. 336–343, Sep. 1996.

[SS98] Samarati, Pierangela; Sweeney, Latanya: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, technical report, SRI International, 1998.

[St16] Steinfeld, Nili: "I agree to the terms and conditions": (How) do users read privacy policies online? An eye-tracking experiment. Computers in Human Behavior, 55:992–1000, 2016.

Ansatz zur Umsetzung von Datenschutz nach der DSGVO im Arbeitsumfeld: Datenschutz durch Nudging

Entwicklung erster Szenarien

Sabrina Schomberg[1], Torben Jan Barev[2], Andreas Janson[3] und Felix Hupfeld[4]

Abstract: Die noch recht neue DSGVO hat einige Änderungen mit sich gebracht, welche sich in der Praxis erst noch bewähren und innovativ umgesetzt werden müssen. Insbesondere das Arbeitsumfeld wird von der fortschreitenden Digitalisierung stark verändert und sieht sich neuen Herausforderungen des Datenschutzes gegenüber. Ein Ansatz, diesen Herausforderungen zu begegnen, könnte die Integration von Privacy Nudges in digitale betriebliche Systeme sein. Ziel von Privacy Nudges ist es dabei, den Entscheidungsprozess in digitalen Entscheidungsumgebungen gezielt zu mehr Schutz von personenbezogenen Daten und Privatheit zu beeinflussen. Diesem Ansatz nähert sich dieser Beitrag interdisziplinär, durch Erkenntnisse aus der verhaltensökonomischen, informatischen und der rechtswissenschaftlichen Literatur. Schließlich werden verschiedene Szenarien für den Einsatz von Privacy Nudges in digitalen Arbeitssystemen beschrieben und bewertet.

Keywords: Privacy Nudges, Digital Nudging, Datenschutz durch Technik, Datenschutz „by Design“ und „by Default“, DSGVO

1 Einleitung

Die fortschreitende Digitalisierung und Vernetzung verändert unser Arbeitsumfeld und die Art und Weise, wie wir arbeiten. Mit dieser Entwicklung gehen einerseits erhebliche Innovationspotentiale einher. So können signifikante Synergieeffekte entstehen, aber auch flexiblere und effizientere Arbeitsmodelle. Andererseits birgt die Digitalisierung von Arbeitsprozessen jedoch auch Risiken. Nicht nur für sensible Unternehmensdaten, sondern auch für die Privatheit von Mitarbeiterinnen und Mitarbeitern. In der Regel werden große Datenmengen anfallen, welche leicht ausgewertet werden können, so dass die Gefahr eines gläsernen Arbeitnehmers erwächst. Viele Arbeitnehmerinnen und Arbeitnehmer sind für das Thema Datenschutz auch noch nicht hinreichend sensibilisiert und geben unter Umständen unfreiwillig viele Daten preis. Hinzu kommt noch das sog. Privacy Paradox z.B. [KM16], also die Feststellung, dass Nutzerinnen und Nutzer den

[1] Universität Kassel, FG Öffentliches Recht, IT-Recht und Umweltrecht von Prof. Dr. Hornung, Henschelstraße 4 (K33), 34127 Kassel, sabrina.schomberg@uni-kassel.de.
[2 3 4] Universität Kassel, FG Wirtschaftsinformatik von Prof. Dr. Leimeister, Pfannkuchstraße 1 (ITeG), 34121 Kassel, torben.barev@uni-kassel.de; andreas.janson@uni-kassel.de; felix.hupfeld@wi-kassel.de.

doi:10.18420/inf2019_52

Datenschutz abstrakt wertschätzen und sich um ihre Privatheit sorgen, aber dennoch sorglos mit ihren personenbezogenen Daten umgehen.

Daher sind neue Ansätze gefragt, die die Vorteile der Digitalisierung nutzen, zugleich jedoch auch den Datenschutz nach der DSGVO und nationalen Gesetzen vollumfänglich umsetzen. Hier bieten sich sog. Nudging-Konzepte an, um die Umsetzung zu begleiten [JS18]. Das Nudging („Anstupsen“), welches seinen Ursprung in der Verhaltensökonomik hat [TS08], soll das Verhalten von Nutzerinnen und Nutzern in digitalen Umgebungen vorhersehbar dahingehend beeinflussen, dass sie datenschutzfreundlichere Entscheidungen treffen. Dies geschieht jedoch nicht durch verbindliche Anweisungen oder gar Verbote, sondern durch Aufmerksamkeitslenkungen, spielbasierte Motivationen [SJ18], Voreinstellungen oder andere „weiche“ Instrumente, die Anreize zu einem bestimmten Verhalten geben.

Die Datenschutz-Grundverordnung hat einige Änderungen mit sich gebracht, die sich in der Praxis erst noch bewähren müssen. So ist zum Beispiel erstmals der Datenschutz durch Technik (insb. in Art. 25 DSGVO) normiert. Datenschutz durch Technik bedeutet, dass der Datenschutz schon in die Technik „eingebaut“ ist [La19b]. Diese datenschutzfreundliche Technik soll es gar nicht erst ermöglichen, dass mehr als nur die erforderlichen Daten erhoben, verarbeitet oder gespeichert werden [Ba17]. Dieser Ansatz wurde bereits in den 90iger Jahren diskutiert, bisher jedoch nicht explizit festgeschrieben [BG17]. In Deutschland wurde der Datenschutz durch Technik in den § 3a BDSG a.F., das Prinzip der Datenvermeidung und Datensparsamkeit, hineingelesen. Danach waren auch bisher schon Maßnahmen, insbesondere der Anonymisierung und Pseudonymisierung, zu treffen. Da § 3a BDSG a.F. jedoch nicht bußgeldbewährt war, wurde dieser in der Praxis eher stiefmütterlich behandelt [BG17], [La19b].

Art. 25 DSGVO ist gem. Art. 83 Abs. 4 lit. a DSGVO mit hohen Bußgeldern bedroht. Unternehmen haben daher ein großes wirtschaftliches Interesse daran, den Datenschutz durch Technik nach der DSGVO auch wirklich umzusetzen. Da die Norm jedoch wenig konkrete Maßnahmen benennt, bleibt den Verantwortlichen ein weiter Spielraum bei der Umsetzung [La19b]. Das kann Fluch und Segen zugleich sein. Einerseits sind Unternehmen dadurch nicht an starre Vorgaben gebunden und können einen individuellen Weg der Umsetzung finden. Andererseits besteht eine große Unsicherheit darüber, ob auch die Aufsichtsbehörden diese individuellen Maßnahmen als ausreichend erachten werden.

Ziel dieses Beitrags ist es, im Rahmen eines interdisziplinären Ansatzes, an der Schnittstelle von Recht und Informatik, die Umsetzung der neuen Vorgaben des Datenschutzes durch Technik in der DSGVO durch digitales Nudging im Arbeitsumfeld zu beschreiben und anhand ausgewählter Szenarien zu bewerten.

2 Privacy Nudging im digitalen Kontext

2.1 Grundlagen des Nudgings im digitalen Kontext

Der Begriff *Nudge* beschreibt per Definition eine Methode, um „das Verhalten von Menschen zu beeinflussen, ohne dabei auf Verbote und Gebote zurückgreifen oder ökonomische Anreize verändern zu müssen" [TS08]. Nudging im offline Bereich kann demnach eine Vielzahl von Ansätzen beinhalten, um Entscheidungen zu beeinflussen. Was gewählt wird, hängt oft davon ab, wie die Entscheidungen präsentiert werden [WSV16]. Individuen tendieren beispielsweise dazu, voreingestellte Optionen eher anzunehmen als diese zu verändern [ZX16]. Eines der prominentesten Beispiele für die Effektivität von Nudges stellt hierbei die Organspende in Österreich dar, welche zu einer signifikant höheren Anzahl von Organspendern, beispielsweise im Vergleich zu Deutschland, führt. Entscheidend ist, dass in Österreich die Zustimmung zur Organspende vorausgesetzt wird. Nicht-Spender müssen sich demnach bewusst mit der Entscheidung auseinandersetzen und widersprechen. In Deutschland wird die Zustimmung nicht vorausgesetzt und muss, beispielsweise in einem Organspendeausweis, pro-aktiv festhalten werden [JG03]. Die Entscheidung wird demnach durch eine andere Voreinstellung, oder einen sogenannten Nudge, maßgeblich beeinflusst.

Nudging basiert auf dem Prinzip des libertären Paternalismus, um Entscheidungen zu beeinflussen. Dies bedeutet, dass ein Individuum zu jeder Zeit eine Entscheidungsoption frei wählen kann (Liberalismus-Komponente). In seiner Entscheidungsfreiheit ist das Individuum nicht eingeschränkt, da keine der Optionen verboten und auch der wirtschaftliche Anreiz der Alternativen nicht bemerkenswert verändert wird. Das Individuum wird aber zu der Entscheidungsoption genudged, die für dieses den vermeintlich größten Nutzen darstellt (Paternalismus-Komponente) [MLJ18].

Beim *digitalen Nudging* wird dieses Konzept auf den digitalen Raum übertragen und entsprechende Designelemente in der Benutzeroberfläche verwendet, um das Verhalten in digitalen Entscheidungsumgebungen zu steuern. Digitale Entscheidungsumgebungen sind Benutzeroberflächen, die es erfordern, dass Menschen Urteile oder Entscheidungen treffen [SWV18], beispielsweise für welche Kollegen die eigenen Kalendereinträge einsehbar sein sollen. Besonders gefährlich am digitalen Nuding ist dabei die Möglichkeit der Verknüpfung verschiedener Daten, welche leicht zur Überwachung führen kann, und die bessere Möglichkeit der Personalisierung von Nudges, welche subtile und effektive Manipulation ermöglichen kann [Sa17].

Eine Unterform der digitalen Nudges sind hierbei die sogenannten *Privacy Nudges*. Privacy Nudging beschreibt eine gezielte Beeinflussung des Entscheidungsprozesses, um Menschen dazu zu bringen, dass diese „bessere“ Entscheidungen in Bezug auf deren Privatheit treffen und gleichzeitig ihre informationelle Selbstbestimmung berücksichtigen [Ac17]. Dies birgt jedoch durchaus auch die Gefahr, das Gegenteil zu bewirken und die Privatsphäre der Nutzer zu verletzen. Einerseits, weil es einfacher ist, effektive per-

sonalisierte Nudges zu gestalten, wenn viele persönliche Daten und Verhaltensmuster bekannt sind. Andererseits, weil sich manche Menschen vielleicht auch bewusst dazu entscheiden so viel wie möglich über sich selbst preiszugeben [SK18].

2.2 Grundlegende Prinzipien des Privacy Nudgings

Untersuchungen haben gezeigt, dass insbesondere Nutzer digitaler Systeme aufgrund kognitiver, emotionaler und sozialer Faktoren oft irrational handeln [Ac17], [TSB10]. Dies lässt sich durch die von *Kahnemann* bekannt gewordene Dualprozesstheorie erklären, die besagt, dass Individuen zwei Denksysteme verwenden. Zwei Systeme sind demnach notwendig, um in der heutigen (digitalen) Arbeitswelt den Überfluss an Informationen besser auswerten zu können und gezielt Entscheidungen zu fällen. System 1 stellt hierbei unsere Intuitionen oder unseren unbewussten Autopiloten dar. System 2 hingegen äußert sich durch unser bewusstes Planen und Kontrollieren. Dies erfordert jedoch deutlich mehr mentale Anstrengung und Zeit. Beide Systeme sind gleichzeitig aktiv und arbeiten meist reibungslos zusammen [Ka13], [Ka03]. Im Arbeitsalltag haben die Individuen hingegen selten genügend Zeit und Informationen, um alle Alternativen vollständig zu bewerten. Anstatt einen systematischen Entscheidungsprozess auszuüben, neigen Individuen dazu, auf so genannte Heuristiken (mentale Abkürzungen) zurückzugreifen [HG17]. Heuristiken sind informelle Faustregeln, die die Komplexität der Urteilsfindung reduzieren und somit Abkürzungen in der Entscheidungsfindung darstellen. Heuristiken sind zwar ein effizienter Weg, um wiederkehrende Aufgaben zu lösen, können aber zu systematischen Fehlern wie Verzerrungen in der Informationsbewertung (Biases) führen [Ka13]. So werden beispielsweise personenbezogene Daten oftmals sorglos offengelegt, da das Risiko der unerwünschten Überwachung mental weniger präsent ist (Verfügbarkeitsheuristik). Diese Fehlschlüsse bedeuten nicht, dass das Verhalten von Individuen unberechenbar und irrational ist. Es ist vielmehr eine systematische und vorhersehbare Abweichung vom rationalen Verhalten. An diesem Punkt kommen Privacy Nudges ins Spiel. Privacy Nudges können beide Denksysteme beeinflussen, indem sie Heuristiken ausnutzen oder ihnen entgegenwirken, um Individuen zu ihrer informationellen Selbstbestimmung zu leiten [WSV16].

3 Rechtliche Rahmenbedingung

Im Kontext des Datenschutzes durch digitales Nudging im Arbeitsumfeld ist insbesondere der Datenschutz durch Technik und der Beschäftigtendatenschutz zu beachten. Darüber hinaus müssen natürlich immer auch die allgemeinen Anforderungen der DSGVO eingehalten werden; insbesondere die in Art. 5 DSGVO erstmals kodifizierten Datenschutzgrundsätze. Diese Datenschutzgrundsätze enthalten jedoch eine Reihe unbestimmter Rechtsbegriffe und legen daher lediglich allgemeine Leitlinien fest, welche dann in den weiteren Normen der DSGVO konkretisiert werden [La19a].

3.1 Datenschutz durch Technik

Mit dem Datenschutz durch Technik muss sich der Verantwortliche möglichst früh auseinandersetzen. Bereits im Entwicklungsstadium sollten einige technische Anforderungen beachtet werden, um die personenbezogenen Daten und die Privatsphäre der Nutzerinnen und Nutzer angemessen zu schützen [Ha18].

Alle Maßnahmen der datenschutzfreundlichen Technikgestaltung gem. Art. 25 Abs. 1 DSGVO sind unter Berücksichtigung des Stands der Technik, der Implementierungskosten und der Art, des Umfangs, der Umstände und der Zwecke der Verarbeitung sowie der unterschiedlichen Eintrittswahrscheinlichkeit und Schwere der mit der Verarbeitung verbundenen Risiken für die Rechte und Freiheiten natürlicher Personen auszuwählen und zu treffen. Dies ist Ausdruck des risikobasierten Ansatzes der DSGVO und begrenzt die Auswahl geeigneter technischer Maßnahmen [BH17]. Es bedarf also nicht immer der theoretisch optimalen Maßnahme, sondern bei geringem Risiko oder besonders hohen Implementierungskosten kann im Einzelfall ggf. auch ein geringerer Schutz ausreichend sein. Daher ist eine Verhältnismäßigkeitsabwägung vorzunehmen, welche im Sinne einer allgemeinen Risiko- und Folgenabschätzung dokumentiert werden sollte, um der Rechenschaftspflicht des Art. 5 Abs. 2, Art. 24 Abs. 1 DSGVO Genüge zu tun [La19b].

Datenschutzfreundliche Voreinstellungen gem. Art. 25 Abs. 2 DSGVO werden wiederum von Teilen der Literatur als eine Konkretisierung der datenschutzfreundlichen Technikgestaltung gem. Art. 25 Abs. 1 DSGVO verstanden, so z.B. [BH17]. Verarbeitungssysteme müssen danach so eingestellt sein, dass nur die für den Zweck der Verarbeitung erforderlichen Daten verarbeitet werden. Es reicht dabei nicht aus, dass der Nutzer eine Wahl- oder Gestaltungsmöglichkeit hat. Personenbezogene Daten dürfen nicht ohne Kenntnis und ohne Zustimmung des Betroffenen verarbeitet werden [Ri18]. Die Zulässigkeit einer Systemeinstellung beurteilt sich also danach, ob die Verarbeitung hinsichtlich Menge, Umfang, Speicherfrist und Zugänglichkeit der personenbezogenen Daten für den Zweck erforderlich ist [Ri18].

Martini war der erste Autor, der im Kontext des Art. 25 Abs. 2 DSGVO das Wort „Nudging“ verwendete [Ma17]. Er geht allerdings von „Nudging mit umgekehrter Stoßrichtung“ durch die DSGVO aus, da die Dienstanbieter nun ihre eigenen wirtschaftlichen Interessen dem Gebot der Datenminimierung unterordnen müssen [Ma18]. *Thaler* und *Sunstein* können aber viel eher so verstanden werden, dass Art. 25 Abs. 2 DSGVO genau die Intention der Autoren von „Nudge“ trifft. Denn es geht um „bessere“ Entscheidungen für den Nutzer und nicht für den Dienstanbieter; also den Angestupsten und nicht den Entscheidungsarchitekten [TS08]. Dass die Dienstanbieter Voreinstellungen datenschutzfreundlich und nicht maximal vorteilhaft für die eigene Gewinnerzielung ausgestalten, wäre daher mutmaßlich auch im Sinne von *Thaler* und *Sunstein*.

Privacy Nudges in Form von datenschutzfreundlichen Voreinstellungen sind mithin in Art. 25 Abs. 2 DSGVO ausdrücklich vorgesehen [HB17]. Fraglich ist, ob darunter auch weitere digitale Nudges gefasst werden können. Bei Auslegung des Wortlauts von Art. 25 Abs. 2 DSGVO dürfte es schwer sein neben Default Nudges, also Voreinstellungen,

auch weitere Arten von Nudges unter diesen zu subsumieren. Weitere Nudges könnten als technische und organisatorische Maßnahmen jedoch unter Art. 25 Abs. 1 DSGVO zu fassen sein. Der Wortlaut des Abs. 1 ist weiter und so unkonkret, dass er durch die Verantwortlichen ausgestaltet werden muss. Eine mögliche Ausgestaltung könnte die Verwendung datenschutzfreundlicher Nudges sein. Dies passt auch insoweit in die Systematik, als dass Art. 25 Abs. 2 DSGVO eine Konkretisierung des Absatz 1 ist (s.o.). Desweitern dürfte dies im Sinne des Verordnungsgebers sein, sofern so personenbezogene Daten geschützt werden können, ohne die Betroffenen ihrer Entscheidungsfreiheit zu berauben (vgl. Art. 1 Abs. 2 DSGVO).

Nach Art. 25 Abs. 3 DSGVO kann eine erfolgreiche Zertifizierung im Sinne des Art. 42 DSGVO oder die Einhaltung genehmigter Verfahrensregeln (Art. 40 Abs. 2 lit. h DSGVO) als ein Faktor herangezogen werden, um die Erfüllung der Anforderungen der Norm nachzuweisen [La19b].

Adressat des Art. 25 DSGVO ist ausdrücklich nur der Verantwortliche, nicht jedoch der Hersteller von Verarbeitungstechnik. Für den Hersteller besteht daher grundsätzlich keine Pflicht zur datenschutzfreundlichen Ausgestaltung seiner Produkte. Er wird lediglich durch Erwägungsgrund 78 dazu „ermutigt“ [BH17]. Da für die Verantwortlichen jedoch eine hohe Strafe droht, werden sie nur solche Produkte kaufen, die den Anforderungen der DSGVO gerecht werden. Deshalb besteht indirekt doch eine Verpflichtung der Hersteller, die typischerweise durch entsprechende Vertragsklauseln umgesetzt werden wird [Ha18].

Der Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen gem. Art. 25 DSGVO wird als Konkretisierung der Pflicht zur Umsetzung technischer und organisatorischer Maßnahmen durch den Verantwortlichen gem. Art. 24 DSGVO verstanden [Ma18]. Art. 25, Art. 32 und Art. 35 DSGVO sind so eng verzahnt, dass sich eine gemeinsame Bearbeitung der verschiedenen Schritte und Prüfungen anbietet. Art. 25 und Art. 32 DSGVO ähneln sich schon vom Wortlaut so sehr, dass eine klare Differenzierung zwischen Maßnahmen gem. Art. 25 DSGVO und Maßnahmen nach Art. 32 DSGVO kaum möglich sein wird. Ohne die Folgenabschätzung gem. Art. 35 DSGVO wiederum wird es kaum möglich sein, die Risiken, die mit der Datenverarbeitung einhergehen, abzuschätzen und dementsprechende Maßnahmen zu ergreifen [Ha18].

3.2 Beschäftigtendatenschutz

Seit dem 25.05.2018 gilt die DSGVO als Verordnung unmittelbar und muss, im Gegensatz zu einer Richtlinie, nicht durch den nationalen Gesetzgeber umgesetzt werden. Sie genießt einen Anwendungsvorrang gegenüber nationalen Regelungen. Es gibt jedoch in der DSGVO eine Vielzahl von Öffnungsklauseln, welche den Mitgliedstaaten wiederum Raum für nationale Regelungen gewähren [KM16]. Eine dieser Öffnungsklauseln ist Art. 88 DSGVO, welcher es den Mitgliedstaaten erlaubt, „spezifischere Vorschriften zur Gewährleistung des Schutzes der Rechte und Freiheiten hinsichtlich der Verarbeitung

von personenbezogenen Beschäftigtendaten im Beschäftigungskontext“ zu erlassen. Durch diesen Wortlaut wird indiziert, dass keine wesentlichen inhaltlichen Abweichungen von den allgemeinen Vorgaben der DSGVO erlaubt sind [Wy17], [TR16], [Ko17]. Art. 88 Abs. 2 DSGVO schreibt vor, dass die nationalen Vorschriften „geeignete und besondere Maßnahmen zur Wahrung der menschlichen Würde, der berechtigten Interessen und der Grundrechte der betroffenen Personen“ umfassen.

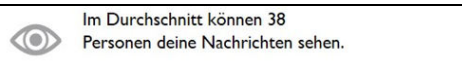
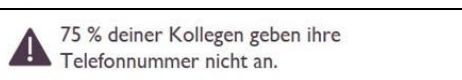
Der deutsche Gesetzgeber hat davon in § 26 BDSG Gebrauch gemacht und orientierte sich dabei erkennbar an § 32 BDSG a.F., welcher zuvor den Beschäftigtendatenschutz regelte [Wy17]. So wurde der Kern der alten Regelung übernommen und es werden nach wie vor alle drei Phasen des Beschäftigungsverhältnisses, nämlich die Begründung, dessen Durchführung und dessen Beendigung, erfasst. Diese strukturelle Ähnlichkeit soll für eine gewisse Kontinuität im deutschen Beschäftigtendatenschutz sorgen [Ko18]. Inhaltlich geht die neue deutsche Regelung des Beschäftigtendatenschutzes jedoch deutlich über die bisherige hinaus [Ko17].

§ 26 Abs. 2 BDSG stellt klar, dass Beschäftigte auch weiterhin im Rahmen des Beschäftigungsverhältnisses in die Verarbeitung ihrer personenbezogenen Daten einwilligen können. Dies ergibt sich zudem schon aus Erwägungsgrund 155 der DSGVO und entspricht auch der bisherigen Rechtsprechung des Bundesarbeitsgerichts. Um dem Über-/Unterordnungsverhältnis von Arbeitgebern und Arbeitnehmern Rechnung zu tragen, werden mit Art. 26 Abs. 2 BSDG jedoch erhöhte Anforderungen an die Freiwilligkeit der Einwilligung gestellt. Der Arbeitgeber muss bei der Beurteilung der Freiwilligkeit immer die im Beschäftigungsverhältnis bestehende Abhängigkeit des Beschäftigten berücksichtigen. Von der Freiwilligkeit der Einwilligung ist jedoch auszugehen, wenn für den Beschäftigten ein rechtlicher oder wirtschaftlicher Vorteil erreicht wird oder sofern Arbeitgeber und Beschäftigter gleichgelagerte Interessen verfolgen [Wy17]. Eine Einwilligung der Arbeitnehmer in die Verarbeitung ihrer personenbezogenen Daten mit inkludierten Privacy Nudges dürfte daher auch unproblematisch möglich sein, da in der Regel sowohl der Arbeitgeber als auch der Arbeitnehmer ein Interesse an datenschutzfreundlicher Ausgestaltung der Datenverarbeitung haben. Beschäftigte müssen jedoch über ihr Widerrufsrecht gem. Art. 7 Abs. 3 DSGVO aufgeklärt werden.

Problematischer könnte allerdings die unklare Rolle des Betriebsrats sein. Weder die DSGVO noch § 26 BDSG befassen sich mit der Frage, ob der Betriebsrat eigenständiger Datenverarbeiter oder Teil des Arbeitgebers als der für die Datenverarbeitung Verantwortliche ist [Ko17], [Ko18]. Der Betriebsrat könnte gem. § 87 BetrVG ein Mitbestimmungsrecht bei der Ausgestaltung der Nudges haben. Privacy Nudges dürften für den Betriebsrat jedoch durchaus zustimmungsfähig sein. Bei Beachtung der überschaubaren Besonderheiten des Beschäftigungsdatenschutzes mit Relevanz für Nudging stehen auch Art. 88 DSGVO und § 26 BDSG der Umsetzung der Vorgaben der DSGVO durch Privacy Nudges nicht entgegen.

4 Szenarien für den Einsatz von Privacy Nudges in digitalen Arbeitssystemen

Um Szenarien für Privacy Nudges zu entwickeln, wurden in einer systematischen Literaturrecherche sechs Privacy Nudge Prinzipien identifiziert, welche nachfolgend mit ihren speziellen Biases, Heuristiken und Prinzipien näher betrachtet werden. Zusätzlich werden Szenarien, in denen sie eingesetzt werden können, näher erläutert. Dabei beziehen wir uns insbesondere auf die bestehende Typologisierung von Privacy Nudges nach *Acquisti et al.* [Ac17] und erweitern diese Sichtweise entsprechend um den Kontext digitaler Arbeitssysteme im Betrieb. Zur Veranschaulichung führen wir in Tabelle 1 zu jedem Privacy Nudge ein Beispiel auf. Die Vorgehensweise der systematischen Literaturrecherche orientiert sich an der vorgeschlagenen Methodik von *vom Brocke et al.* [Vo15]. Um die Thematik der Privacy Nudges umfassend zu erfassen, wurde auf Beiträge aus sechs verschiedenen Datenbanken zugegriffen. Neben der AIS eLibray wurden die ACM Digital Library und die IEEE Xplore Digital Library als klassische Repräsentanten der Datenbanken im Bereich Informationssysteme ausgewählt. Das Social Science Research Network (SSRN), ScienceDirect und EBSCOhost wurden hinzugenommen, um auch verhaltensbezogene und psychologische Quellen gezielt zu integrieren. Die deutsche Literatur wurden mit englischsprachigen Beiträgen ergänzt, um den internationalen Stand der Forschung abzubilden.

Privacy Nudge	**Beispiel**
Default	
Farbelemente	
Information	
Feedback	
Zeitverzögerung	
Soziale Norm	

Tab.1: Beispiele digitaler Privacy Nudges im Arbeitsumfeld

Defaults

Default Nudges beschreiben Standardeinstellungen im System. Da Individuen in digitalen Umgebungen die Privatsphäre-Einstellungen häufig nicht ihren Bedürfnissen anpassen, bleibt die voreingestellte Option (der Status-quo) übermäßig bevorzugt und meist unverändert (Status-quo Bias) [Ac17], [TS08]. Zudem wird die voreingestellte Option als Referenzpunkt für das Abwägen der Entscheidungsoptionen herangezogen. Dieser „Anker" wird von Individuen unbewusst wahrgenommen. Jede Entscheidungsoption wird nun gegen diese Alternative abgewägt und das Entscheidungsverhalten in diese Richtung beeinflusst [TK74]. *Hummel* und *Maedche* bewerten Defaults tendenziell als die stärksten Nudges [HM19]. In Bezug auf Privacy Nudging gelten Defaults als sehr effektiv, da sie in digitalen Arbeitssystemen standardmäßig das Maß der Datensparsamkeit vorgeben [Ac17].

Farbelemente

Auch Farbelemente können als Privacy Nudges verwendet werden. Farbliche Hinterlegungen lenken hierbei die Aufmerksamkeit auf ausgewählte Elemente, um bestimmte Entscheidungsalternativen verstärkt hervorzuheben. Bei mobilen Apps kann beispielsweise die Schaltfläche zur Datenfreigabe farblich stärker betont werden. Im aufgezeigten Beispiel wird der „privat"-Button in grüner Farbe markiert und Individuen dazu angehalten, diese Option zu wählen. Im Rahmen der digitalen Arbeit wären sensible Daten nun ausschließlich für eine bestimmte Zielgruppe oder nur für das Individuum selbst zugänglich [Al15]. Die Vorteile der Farbelemente zeigen sich vor allem in der einfachen Umsetzung solcher Nudges, die das Individuum schnell und effektiv dazu bewegen, seine Entscheidungen bezüglich des Datenschutzes und der Privatsphäre zu überdenken.

Information

Die Wahrscheinlichkeit einer Verletzung der Privatsphäre ist häufig für Individuen nicht nachvollziehbar und wird oft unterschätzt. Das Individuum tendiert dann dazu, am Arbeitsplatz risikoreiche Entscheidungen in Bezug auf den Schutz der eigenen Privatsphäre zu treffen. Dies lässt sich unter anderem auf die Repräsentationsheuristik zurückführen, bei der Individuen dazu tendieren, die Häufigkeit der Beobachtungen eines Ereignisses fälschlicherweise mit dessen Eintrittswahrscheinlichkeit in Verbindung zu bringen. Auch die Verfügbarkeitsheuristik spielt hierbei eine große Rolle, bei der Entscheidungen auf Informationen begründet werden, die mental leicht verfügbar sind [Ac17], [TK74]. Um diesen Heuristiken entgegenzuwirken, wird das Individuum über Risiken und Konsequenzen seines Handelns aufgeklärt. Basierend auf diesen Informationen kann das Individuum eine fundierte Entscheidung in Bezug auf die eigene Privatsphäre treffen [Ac17].

Feedback

Einen weiteren Privacy Nudge stellt die Bereitstellung von Feedback dar, welches auf das bisherige Nutzungsverhalten einer Person hinweist. Dies schafft beim Individuum ein Bewusstsein über seine bisherigen und aktuellen Entscheidungen und dessen Konse-

quenzen [Ac17]. Ein Beispiel für Privacy Nudging durch Feedback ist ein Fortschritts-Balken, der z.B. beim Registrierungsprozess im Arbeitssystem die Stärke eines Passworts illustriert oder die Menge der eingegebenen Daten im Profil widerspiegelt. So werden Individuen spielerisch dazu angehalten, ein komplexeres Passwort zu wählen bzw. weniger Daten im System zu hinterlegen.
Entscheidend für ein erfolgreiches Nudging durch Feedback ist die Art und Weise der Darstellung. Insbesondere Textbenachrichtigungen ohne Ton, die den Arbeitsfluss nicht einschränken, gelten als effektiv [Mi17].

Zeitverzögerung

Bei digitalen Entscheidungen über die Privatsphäre werden oftmals risikoreiche und wenig durchdachte Entscheidungen ohne Anbetracht der möglichen Spätfolgen begünstigt. Dem zugrunde liegt das sogenannte Hyperbolic Discounting, bei dem der unmittelbare Nutzen überschätzt und später eintretende Kosten unterschätzt werden [Ac17]. Um diesem entgegenzuwirken, kann eine zeitliche Verzögerung als Privacy Nudge verwendet werden [Wa14]. Beispielsweise wird ein Countdown von fünf Sekunden eingesetzt, bevor eine Nachricht mit riskanten Inhalten im Firmennetzwerk veröffentlicht wird. In diesen Sekunden besteht weiterhin die Möglichkeit, die Nachricht zurückzuziehen, zu bearbeiten oder die Wartezeit direkt zu überspringen. So soll das Individuum dazu bewegt werden, weniger impulsiv zu agieren sowie die Nachricht und mögliche negative Konsequenzen zu überdenken [Ac17]. Während die Zeitverzögerung große Effektivität verspricht, sollte beim Einsatz dieses Privacy Nudges bedacht werden, dass die Verzögerung der Aktion auch als störend empfunden werden kann.

Soziale Norm

Die Wirkung dieses Privacy Nudges basiert auf dem Prinzip der sozialen Normen. Das Individuum leitet dabei aus dem Verhalten seiner Mitmenschen ab, inwiefern es angemessen ist, persönliche Informationen zu teilen [Ch16], [Co16]. Beispielsweise ist für das Individuum erkenntlich, dass 75 % Prozent der Kollegen die eigene Telefonnummer nicht im Arbeitsprofil angegeben haben. Diese Information wird nun als Referenzpunkt für das eigene Verhalten herangezogen (Ankerheuristik) [Ac17]. Eine Studie im Rahmen der Vergabe von Zugriffsberechtigungen für Smartphone Apps hat dabei gezeigt, dass die soziale Norm auch entgegen einer Datensparsamkeit wirken kann. Falls die Mehrheit den Zugriff einer App auf bestimmte Daten zulässt, könnten Individuen dazu verleitet werden, sich ebenso zu verhalten [ZX16]. Diese Nudges sollten daher mit Bedacht verwendet werden, um Individuen zu besseren Entscheidungen in Bezug auf den Schutz ihrer Daten zu befähigen [Ch16].

5 Abschließende Bewertung

Privacy Nudges können eine effektive Methode darstellen, das Verhalten von Nutzerinnen und Nutzern in digitalen Arbeitsumgebungen vorhersehbar dahingehend zu beein-

flussen, dass sie datenschutzfreundlichere Entscheidungen treffen. Die Wirkung der Privacy Nudges ist dabei stark vom Kontext abhängig. Die vorgestellten Szenarien können bei der Auswahl der richtigen Nudges unterstützen, damit diese ihre volle Wirkung entfalten. Außerdem sollten die Nudges so gewählt werden, dass diese den Arbeitsprozess nicht behindern. Insbesondere bei Nudges, welche auf dem Prinzip der sozialen Norm basieren, ist darauf zu achten, dass diese nicht in die falsche Richtung wirken und die Arbeitnehmer zu mehr Datenoffenlegung verleiten. Generell sollten Nudges möglichst transparent gestaltet sein, um der Gefahr der Manipulation entgegenzuwirken.

Die Personalisierung von Nudges kann zudem deren Effektivität zusätzlich erhöhen [Su15]. Dies könnte daher ein vielversprechender, weitergehender Schritt für Privacy Nudges sein. Dafür ist zu erforschen, wie personalisierte Nudges automatisch umgesetzt werden könnten und welche neuen rechtlichen Fragestellungen sich daraus ergeben. Zudem könnten edukative Nudges, welche zur Reflektion über die Preisgabe von Daten anregen, Lernprozesse befördern und damit Nutzer digitaler Angebote befähigen, bessere Entscheidungen hinsichtlich ihrer Privatsphäre zu treffen.

Bei den Default Nudges in digitalen Arbeitssystemen handelt es sich um eine Maßnahme im Sinne des Art. 25 Abs. 2 DSGVO. Die Literatur legt nahe, dass Default Nudges am effektivsten sind, da sie in digitalen Arbeitssystemen standartmäßig das Maß der Datensparsamkeit vorgeben (vgl. Kapitel 4). Alle anderen vorgestellten Beispiele digitaler Privacy Nudges im Arbeitsumfeld wären nach der hier vertretenen Auffassung (siehe 3.1) rechtlich als Datenschutzmaßnahmen durch Technikgestaltung gem. Art. 25 Abs. 1 DSGVO einzuordnen.

Um die Vorgaben des Art. 25 DSGVO vollständig im Unternehmen umzusetzen und sich nicht der Gefahr eines hohen Bußgeldes gem. Art. 83 Abs. 4 DSGVO auszusetzen werden weitere technische und organisatorische Maßnahmen zu treffen sein. Privacy Nudges können jedoch eine dieser Maßnahme im Sinne des Art. 25 DSGVO sein, um den Schutz von personenbezogenen Daten und der Privatheit von Mitarbeiterinnen und Mitarbeitern zu verbessern. Somit können Privacy Nudges einen wichtigen Beitrag dazu leisten den sehr abstrakten Art. 25 DSGVO mit Leben zu füllen.

Danksagung

Dieser Artikel wurde im Rahmen des Projekts „Nudger“ (www.nudger.de; Förderkennzeichen: 16KIS0890K; 16KIS0891) unter der Projektträgerschaft des VDI/VDE-IT erarbeitet und mit den Mitteln des Bundesministeriums für Bildung und Forschung gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autoren.

Literaturverzeichnis

[Ac17] Acquisti, A. et al.: Nudges for Privacy and Security. In ACM Computing Surveys, 2017, 50; S. 1–41.

[Al15] Almuhimedi, H. et al.: Your Location has been Shared 5,398 Times! In (Begole, B. et al. Hrsg.): Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15. ACM Press, New York, New York, USA, 2015; S. 787–796.

[Ba17] Barlag, C.: § 3 VII. Datenschutz durch Technikgestaltung. In (Roßnagel, A. Hrsg.): Europäische Datenschutz-Grundverordnung. Vorrang des Unionsrechts - Anwendbarkeit des nationalen Rechts. Nomos, Baden-Baden, 2017.

[BG17] Baumgartner, U.; Gausling, T.: Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen. Was Unternehmen jetzt nach der DS-GVO beachten müssen. In ZD, 2017; S. 308–313.

[BH17] Bieker, F.; Hansen, M.: Datenschutz "by Design" und "by Default" nach der neuen europäischen Datenschutz-Grundverordnung. In RDV, 2017; S. 165–170.

[Ch16] Chang, D. et al.: Engineering Information Disclosure: CHI'16 Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems; S. 587–597.

[Co16] Coventry, L. M. et al.: Personality and Social Framing in Privacy Decision-Making: A Study on Cookie Acceptance. In Frontiers in psychology, 2016, 7; S. 1341.

[Ha18] Hartung, J.: Art. 25. In (Kühling, J.; Buchner, B. Hrsg.): Datenschutz-Grundverordnung/BDSG. Kommentar. C.H. Beck, München, 2018.

[HB17] Herfurth, C.; Benner-Tischler, A.: Nudging in der DS-GVO und die Wirkung von Privacy by Default. In ZD-Aktuell, 2017.

[HG17] Hertwig, R.; Grüne-Yanoff, T.: Nudging and Boosting: Steering or Empowering Good Decisions. In Perspectives on psychological science a journal of the Association for Psychological Science, 2017, 12; S. 973–986.

[HM19] Hummel, D.; Maedche, A.: How effective is nudging? A quantitative review on the effect sizes and limits of empirical nudging studies. In Journal of Behavioral and Experimental Economics, 2019, 80; S. 47–58.

[JG03] Johnson, E. J.; Goldstein, D.: Medicine. Do defaults save lives? In Science (New York, N.Y.), 2003, 302; S. 1338–1339.

[JS18] Janson, A.; Schöbel, S.: Nudging Privacy in Digital Work Systems. Towards the Development of a Desgin Theory. In International Conference on Information Systems (ICIS), 2018.

[Ka03] Kahneman, D.: Maps of Bounded Rationality: Psychology for Behavioral Economics. In American Economic Review, 2003, 93; S. 1449–1475.

[Ka13] Kahneman, D.: Thinking, fast and slow. Farrar, New York, 2013.

[KM16] Kühling, J.; Martini, M.: Die Datenschutz-Grundverordnung: Revolution oder Evolution im europäischen und deutschen Detenschutzrecht? In EuZW, 2016; S. 448–453.

[Ko17] Kort, M.: Der Beschäftigtendatenschutz gem. § 26 BDSG-neu. Ist die Ausfüllung der Öffnungsklausel des Art. 88 DS-GVO geglückt? In ZD, 2017; S. 319–323.

[Ko18] Kort, M.: Die Bedeutung der neueren arbeitsrechtlichen Rechtsprechung für das Verständnis des neuen Beschäftigtendatenschutzes. In NZA, 2018; S. 1097–1105.

[La19a] Laue, P.: § 1. Einführung. In (Laue, P.; Kremer, S. Hrsg.): Das neue Datenschutzrecht in der betrieblichen Praxis. Nomos, Baden-Baden, 2019.

[La19b] Laue, P.: § 7. Technischer und organisatorischer Datenschutz. In (Laue, P.; Kremer, S. Hrsg.): Das neue Datenschutzrecht in der betrieblichen Praxis. Nomos, Baden-Baden, 2019.

[Ma17] Martini, M.: Art. 25. In (Paal, B. P.; Pauly, D. A. Hrsg.): Datenschutz-Grundverordnung. C.H.Beck, München, 2017.

[Ma18] Martini, M.: Art. 25. In (Paal, B. P.; Pauly, D. A. Hrsg.): Datenschutz-Grundverordnung, Bundesdatenschutzgesetz. C.H. Beck, München, 2018.

[Mi17] Micallef, N. et al.: Stop annoying me! In (Soro, A. et al. Hrsg.): Proceedings of the 29th Australian Conference on Computer-Human Interaction - OZCHI '17. ACM Press, New York, New York, USA, 2017; S. 371–375.

[MLJ18] Tobias Mirsch, Christiane Lehrer, and Reinhard Jung: Making Digital Nudging Applicable: The Digital Nudge Design Method: Thirty Ninth International Conference on Information Systems, San Francisco.

[Ri18] Richter, P.: Datenschutz durch Technik und datenschutzfreundliche Voreinstellung. In (Jandt, S.; Steidle, R. Hrsg.): Datenschutz im Internet. Rechtshandbuch zu DSGVO und BDSG. Nomos, Baden-Baden, 2018; S. 356–374.

[Sa17] Sascha Lobo: Nudging - Du willst es doch auch. Oder? In Spiegel Online, 2017.

[SJ18] Schöbel, S.; Janson, A.: Is it All About Having Fun? - Developing a Taxonomy to Gamifiy Information Systems. In ECIS 2018 Proceedings, 2018.

[SK18] Sandfuchs, B.; Kapsner, A.: Privacy Nudges: Conceptual and Constitutional Problems. In (Bürk, S. et al. Hrsg.): Privatheit in der digitalen Gesellschaft. Duncker & Humblot, Berlin, 2018; S. 319–338.

[Su15] Sunstein, C. R.: Do People Like Nudges? In SSRN Electronic Journal, 2015.

[SWV18] Schneider, C.; Weinmann, M.; Vom Brocke, J.: Digital nudging. In Communications of the ACM, 2018, 61; S. 67–73.

[TK74] Tversky, A.; Kahneman, D.: Judgment under Uncertainty: Heuristics and Biases. In Science (New York, N.Y.), 1974, 185; S. 1124–1131.

[TR16] Taeger, J.; Rose, E.: Zum Stand des deutschen und europäischen Beschäftigtendatenschutzes. In BB (Betriebs-Berater), 2016; S. 819–831.

[TS08] Thaler, R. H.; Sunstein, C. R.: Nudge. Improving decisions about health, wealth, and happiness. Yale University Press, New Haven, 2008.

[TSB10] Thaler, R. H.; Sunstein, C. R.; Balz, J. P.: Choice Architecture. In SSRN Electronic Journal, 2010.

[Wa14] Wang, Y. et al.: A field trial of privacy nudges for facebook. In (Jones, M. et al. Hrsg.): Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14. ACM Press, New York, New York, USA, 2014; S. 2367–2376.

[WSV16] Weinmann, M.; Schneider, C.; Vom Brocke, J.: Digital Nudging. In Business & Information Systems Engineering, 2016, 58; S. 433–436.

[Wy17] Wybitul, T.: Der neue Beschäftigtendatenschutz nach Art. 26 BDSG und Art. 88 DSGVO. In NZA, 2017; S. 413–419.

[ZX16] Zhang, B.; Xu, H.: Privacy Nudges for Mobile Applications: Effects on the Creepiness Emotion and Privacy Attitudes. In (Gergle, D. et al. Hrsg.): Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing - CSCW '16. ACM Press, New York, New York, USA, 2016; S. 1674–1688.

Herausforderungen für die Anonymisierung von Daten*

Christian Winter[1] Verena Battis[1] Oren Halvani[1]

Abstract: Unternehmen, Wissenschaftler und staatliche Stellen haben ein großes Interesse, neue Erkenntnisse aus Daten zu gewinnen. Dabei müssen Datenschutzregeln eingehalten werden. Anonymisierung ist auf den ersten Blick eine attraktive Lösung, um Datenschutz und Analyseinteressen miteinander zu vereinbaren. Jedoch ist eine korrekte Anonymisierung, die jeglichen Personenbezug entfernt, kaum zu erreichen und schwerlich zu garantieren, wenn gleichzeitig möglichst viel des Informationsgehalts der Daten erhalten werden soll. Wir geben in diesem Aufsatz einen Überblick über den Stand der Technik der Anonymisierung für strukturierte und unstrukturierte Daten, arbeiten die bestehenden Defizite heraus und formulieren Herausforderungen, die auf dem Weg zu besseren Anonymisierungsverfahren gelöst werden müssen.

Keywords: Anonymisierung; personenbezogene Daten; strukturierte Daten; unstrukturierte Daten; Textdaten; maschinelles Lernen; Informationsverlust

1 Motivation

In der heutigen Zeit werden Daten als das neue Gold oder Öl angesehen. In vielen Szenarien geht es um *Daten über natürliche Personen*, etwa um deren Surfverhalten im Internet, um deren Konsumverhalten, um Bewegungsprofile von Personen, um deren finanzielle oder gesundheitliche Situation und Historie, um die Interessen, Gesinnungen und Kontakte von Personen oder um öffentliche oder private Kommunikation. Menschen haben jedoch generell ein Bedürfnis nach *Privatsphäre* und auch ein Grundrecht (Artikel 7 der EU-Grundrechtecharta) darauf. In Bezug auf Datenverarbeitung wird das Recht auf Privatsphäre durch das Grundrecht auf *Datenschutz* (Artikel 8 der EU-Grundrechtecharta) ergänzt, welches durch die Datenschutzgrundverordnung (DSGVO) konkretisiert ist.

Auf der einen Seite gibt es also das Interesse, viele Daten zu sammeln, zusammenzuführen und auszuwerten. Unter dem Leitmotiv *Big Data* sind die technischen Möglichkeiten und die Erwartungen diesbezüglich stark gestiegen. Insbesondere Verfahren des maschinellen Lernens versprechen gesteigerte Möglichkeiten für Rückschlüsse aus Daten. Auf der anderen Seite gibt es die *betroffenen Personen* mit den Grundrechten auf Privatsphäre und Datenschutz. Diese Personen sind verschiedenen Risiken durch die Datenverarbeitung ausgesetzt, beispielsweise, dass sie durch das Gefühl ständiger Beobachtung ihre Handlungsfreiheit

* Diese Arbeit wurde vom Hessischen Ministerium des Innern und für Sport im Teilprojekt „Privacy und Big Data“ im Rahmen des Projekts „Cybersicherheit für die digitale Verwaltung“ gefördert.

[1] Fraunhofer-Institut für Sichere Informationstechnologie SIT, Rheinstraße 75, 64295 Darmstadt
{vorname}.{nachname}@sit.fraunhofer.de

doi:10.18420/inf2019_53

einschränken, dass sie durch automatisierte Entscheidungen diskriminiert werden oder dass sie durch Fehler in Daten und Systemen und den daraus resultierenden Fehlentscheidungen signifikant benachteiligt werden.

Der Datenschutz dient dazu, die Risiken für Betroffene zu minimieren und rechtliche Mittel zu schaffen, um das Machtverhältnis zwischen Datennutzern und Betroffenen auszugleichen. Von vielen Datennutzern wird der Datenschutz jedoch als Hindernis wahrgenommen, welches möglicherweise sogar manche Vorhaben zur Gewinnung von operativen oder grundsätzlichen Erkenntnissen oder zur Entwicklung und Erprobung neuer Verfahren vereitelt. Dieser Interessenskonflikt muss von Einzelfall zu Einzelfall bewertet und gelöst werden. Generell kann man jedoch unterscheiden, ob eine Datenverarbeitung auf konkrete Personen abzielt, etwa im Endkundengeschäft eines Unternehmens oder bei der Auslieferung von Werbung, oder ob es nur um Erkenntnisse über größere Personengruppen geht, z. B. statistische Eigenschaften, Zusammenhänge und Tendenzen, etwa in der Geschäftsanalytik, in Testumgebungen oder in der Forschung. In der zweiten Konstellation ist die *Anonymisierung* von Daten ein probates Mittel zur Ermöglichung der Datennutzung, da anonymisierte Daten keinen Personenbezug mehr enthalten und somit nicht mehr dem Datenschutz unterliegen.

In den nachfolgenden Abschnitten werden die bestehenden Herausforderungen für die Anonymisierung detailliert herausgearbeitet. Dabei werden primär technische Herausforderungen betrachtet, aber in Abschnitt 2 wird deutlich, dass es auch bei den rechtlichen Rahmenbedingungen Herausforderungen gibt. Technische Anonymisierungsverfahren müssen nach der Art der vorliegenden Daten gewählt werden. Wegen der langen Tradition der Verarbeitung und Anonymisierung von strukturierten Daten untersuchen wir zunächst dieses Gebiet (s. Abschnitt 3). Da viele Information jedoch nicht strukturiert, sondern als Fließtexte vorliegen und zunehmend auch bei automatischen Analysen einbezogen werden, betrachten wir auch Textdaten (s. Abschnitt 4). Aufgrund der zunehmenden Relevanz von maschinellem Lernen gehen wir auch auf die hierdurch begründeten besonderen Herausforderungen für die Anonymisierung ein (s. Abschnitt 5). Nicht genauer betrachtet wird in dieser Publikation die Anonymisierung von Multimediadaten (Bild, Audio, Video) aufgrund der Weite dieses Feldes, welches eine eigenständige Arbeit erfordern würde.

2 Definition und Überprüfbarkeit von Personenbezug und Anonymität

Der Datenschutz regelt den Umgang mit *personenbezogenen Daten*. Grundsätzlich besteht jedoch eine große Schwierigkeit in einer exakten Definition solcher Daten. Analog dazu ist es schwierig zu definieren, wann Daten *nicht personenbezogen*, also *anonym*, sind.

Die DSGVO definiert personenbezogene Daten als solche, „die sich auf eine identifizierte oder identifizierbare natürliche Person [. . .] beziehen“ (Artikel 4 Nr. 1 DSGVO). Während die Identifizierbarkeit von Personen weiter erläutert wird, bleibt unspezifiziert, wie konkret das Beziehen auf eine natürlichen Person sein muss oder wie vage es sein kann, damit ein Personenbezug im Sinne des Gesetzes gegeben ist. Etwas spezifischer in diesem Aspekt

ist die alte Fassung des Bundesdatenschutzgesetzes (BDSGaF), welche personenbezogene Daten als „Einzelangaben“ zu natürlichen Personen definiert (§ 3 Abs. 1 BDSGaF). Hier wird deutlich, dass es nicht um allgemeine statistische Aussagen über Personen geht, sondern um Angaben über einzelne, konkrete Personen. Es ist jedoch zu berücksichtigen, dass in vielen Fällen auch bei Mehrpersonenangaben *Rückschlüsse* über einzelne der einbezogenen Personen getroffen werden können. Dadurch ist der Übergang zwischen Einzelangaben und anonymen Statistiken fließend und es ist nicht klar, wo die Grenze im Sinne des BDSGaF liegt und noch weniger, wo sie im Sinne der DSGVO liegt.

Da in der Datenschutzrichtlinie, welche durch die DSGVO abgelöst wurde, die Definition von personenbezogenen Daten im Kern identisch mit der Definition aus der DSGVO ist, sind die Deutungen der ehemaligen Artikel-29-Datenschutzgruppe zu dem Begriff der personenbezogenen Daten auch im Kontext der DSGVO relevant. In der Opinion 05/2014 [DSG14] befasst sich die Gruppe mit dem Thema Anonymisierung und hier werden Rückschlüsse als eines der zentralen Risiken betrachtet. Dieses Risiko wird folgendermaßen charakterisiert: „Inference, which is the possibility to deduce, with significant probability, the value of an attribute from the values of a set of other attributes“, wonach eine sehr weitreichende Definition von Rückschlüssen gewählt wird. Demnach kann der Personenbezug von Daten auch weit in allgemeine statistische Aussagen hinein bestehen. Der Interpretationsspielraum beim Begriff des Personen*bezugs* muss ausgeräumt werden, um das Problem der klaren Abgrenzung von personenbezogenen und anonymen Daten zu lösen. Hier ist insbesondere die Rechtsentwicklung gefragt.

Neben einer präzisen, formalen Definition von personenbezogenen Daten fehlt es auch an *Kriterien*, mit denen Daten zweifelsfrei überprüft werden können, ob ein Personenbezug vorhanden ist oder ob die Daten anonym sind. Mangels einer solchen Überprüfbarkeit gibt es keine *Garantie*, dass ein nach dem Stand der Technik anonymisierter Datenbestand auch tatsächlich anonym ist. Viele der bestehenden Anonymisierungsverfahren für strukturierte Daten (s. Abschnitt 3) arbeiten zwar mit gewissen Anonymitätsmaßen als formale Kriterien, aber diese Maße geben eher einen Grad von Anonymität unter bestimmten, teils impliziten, Annahmen über die Daten und über mögliche Angriffe auf die Anonymität wieder. Hier stellt sich zum einen die Frage, welcher Anonymitätsgrad nach dem jeweiligen Maß ausreichend ist, um von Anonymität im Sinne eines nicht mehr vorhandenen Personenbezugs sprechen zu können, und zum anderen, ob nicht auch jenseits der von dem verwendeten Maß betrachteten Aspekte Angriffe auf die Anonymität möglich sind. Für unstrukturierte Daten fehlen Anonymitätskriterien sogar gänzlich.

3 Anonymisierung strukturierter Daten

Strukturierte Daten werden oft tabellarisch dargestellt, so dass jede Spalte ein bestimmtes Attribut enthält und jede Zeile einen einzelnen Datensatz. Bei personenbezogenen Tabellen ist typischerweise je eine Zeile einer Person zugeordnet. Für die Anonymisierung solcher Daten gibt es verschiedene elementare *Strategien*:

Generalisierung: Die jeweiligen Attributwerte werden durch weniger genaue Angaben ersetzt, etwa durch Intervalle bei numerischen Daten oder durch übergeordnete Kategorien bei kategorischen Daten.

Löschung: Der Inhalt einzelner Zellen, Spalten oder Zeilen wird gelöscht. Dies entspricht einer Generalisierung zu einem allumfassenden und nichtssagenden Wert, etwa „*“.

Mikroaggregierung: Die Daten werden nach Ähnlichkeit in den Attributwerten gruppiert (engl. *clustering*) und pro Gruppe werden die einzelnen Werte zu einem repräsentative Werte zusammengefasst, etwa dem Mittelwert oder Median.

Verfälschung: Ein Teil der Daten oder alle Daten werden zufällig abgewandelt. Dies kann z. B. dadurch erreicht werden, dass zu den Werten zufällige Störungen hinzugefügt werden, dass verschiedene Einträge in der Tabelle vertauscht werden oder dass eine künstliche Tabelle unter Orientierung an der Originaltabelle synthetisiert wird.

Neben diesen verschiedenen Ansätzen zur Anonymisierung der Daten gibt es auch die Strategie, nicht die Daten in anonymisierter Form herauszugeben, sondern die gewünschte Analyse zu den Originaldaten in geschützter Umgebung zu bringen, die Analyse dort durchzuführen und nur die Ergebnisse vor der Herausgabe zu anonymisieren. Dies kann einfacher sein und präzisere Ergebnisse liefern, aber es muss stets die rechtliche Zulässigkeit einer solchen Verarbeitung geprüft werden. Zudem schwindet der Vorteil und kann sich in das Gegenteil kehren, wenn viele Analysen durchgeführt werden sollen, da bei der Anonymisierung der Ergebnisse dann auch die Querbeziehungen zwischen allen Ergebnissen berücksichtigt werden müssen.

Zum Bestimmen des Anonymitätsgrades von Daten, die mit den oben genannten Strategien behandelt worden sind, gibt es verschiedene *Kriterien* bzw. *Maße*. Diese Maße unterscheiden sich darin, welche Annahmen über das Hintergrundwissen eines Angreifers und über die Art des zu erreichenden Schutzes gemacht werden. Minimalen Schutz bieten die Kriterien *k-Map* [Sw01] und *δ-Presence* [NAC07], da hier angenommen wird, dass die in der Tabelle erfassten Individuen aus einer größeren Population stammen, ein Angreifer aber nicht wissen kann, ob eine bestimmte Person in der Tabelle enthalten ist. Das bekanntest Anonymitätskriterium ist *k-Anonymität* [Sw01]. Nach diesem Kriterium muss es jeweils mindestens k für eine Person in Frage kommende Einträge in der Tabelle geben, so dass eine Re-Identifikation nicht möglich ist. Da dennoch möglicherweise einzelne Attribute einer Person durch eine k-anonyme Tabelle offengelegt werden können, wurde das Kriterium zu *l-Diversität* [Ma06] und *t-Closeness* [LLV07] weiterentwickelt. Grundsätzlich anders ist das Konzept von *Differential Privacy* [Dw06a]. Hier wird die Anonymität daran gemessen, wie sehr sich das Ergebnis durch Weglassen oder Hinzufügen einer Person ändern kann, und somit wie viel an Information maximal über eine Person offenbart wird.

3.1 Algorithmen für k-Anonymität und verwandte Kriterien

Es gibt eine Vielzahl von *Algorithmen* zum Erreichen der verschiedenen Anonymitätskriterien. Insbesondere für k-Anonymität und die daran anknüpfenden Maße gibt es eine Vielzahl von Algorithmen basierend auf Generalisierung und Löschung. Einfachere Algorithmen beschränken sich auf eine Generalisierung auf der Attribut-Ebene, d. h. es wird für eine Tabellenspalte insgesamt festgelegt, welcher Wert zu welchem generalisiert wird („global recoding"), während komplexere Algorithmen die Generalisierung auf Zell-Ebene festlegen können („local recoding"). Die zweite Gruppe von Algorithmen kann das Ziel mit weniger Informationsverlust erreichen, jedoch ist die Durchführung bei realen Tabellen meist zu aufwändig, da der Aufwand zum Finden einer optimalen Generalisierung bei naiver Suche exponentiell mit der Anzahl der Tabellenzellen steigt. Diese Optimierungsaufgabe ist in der Tat NP-schwer [Du07], so dass keine effizienten Algorithmen existieren. Aber selbst die erste Gruppe von Algorithmen kann bei großen Datentabellen, insbesondere, wenn viele Attribute vorhanden sind, zu aufwändig werden. Algorithmen auf Basis von Mikroaggregierung können eine gute Effizienz aufweisen und gleichzeitig mehr Informationen erhalten als Generalisierungen auf Attribut-Ebene.

Bei allen Verfahren, die k-Anonymität oder verwandte Eigenschaften auf den Daten sicherstellen, ist zu beachten, dass diese Eigenschaften keine Anonymität garantieren (vgl. Abschnitt 2). Sie schützen nur gegen bestimmte Risiken und auch nur, wenn die Annahmen über das Hintergrundwissen der Angreifer und über die Eigenschaften der Daten korrekt sind. Um das Problem von nicht berücksichtigten Angriffsmöglichkeiten zu lösen, muss die Forschung entweder ultimative Anonymitätskriterien finden oder sie muss wenigstens Anwender dabei unterstützen, Schutzlücken oder unpassende Annahmen aufzudecken. Für zweiteres sollten die existierenden Anonymitätskriterien durch formale *Angreifermodelle* ergänzt werden, welche die angenommenen Fähigkeiten und das angenommene (Hintergrund-) Wissen von Angreifern explizit machen. Anwender können damit leichter erkennen, was in ihren Szenarien durch welche Anonymisierung tatsächlich geschützt wird. Zusätzlich muss der Anwender aber auch stets die Semantik der vorhanden Datenattribute bei der Wahl der Anonymisierung beachten.

3.2 Algorithmen für Differential Privacy

Für Differential Privacy gibt es ebenfalls eine Reihe von Algorithmen. Diese Algorithmen nutzen die Strategie der zufälligen Verfälschung. Der *Laplace-Mechanismus* [Dw06b] ist für Frage-Antwort-Systeme geeignet, bei denen nur die aus den Originaldaten gewonnen Antworten in anonymisierter Form herausgegeben werden sollen. Dazu gibt es ein Privacy-Budget, welches nach und nach von den Antworten aufgebraucht wird, so dass sukzessive der Informationsgehalt von Antworten reduziert (d. h. die Störung erhöht) wird oder irgendwann gar keine Antworten mehr gegeben werden können, wenn das Privacy-Budget verbraucht ist. Der *Exponential-Mechanismus* [MT07] hingegen kann genutzt werden, um synthetische

Tabellen nach dem Vorbild der Originaltabelle zu erzeugen [BLR08]. Dabei wird zum einen die Nähe zu den Originaldaten mit höheren Wahrscheinlichkeiten begünstigt und zum anderen wird über einen Parameter gesteuert, wie groß die Streuung ist, um das Kriterium von Differential Privacy zu erfüllen. Der Exponential-Mechanismus ist jedoch mit sehr hohem Aufwand verbunden.

Ein weiteres Verfahren für Differential Privacy sind *randomisierte Antworten*, welche bereits lange in sozialwissenschaftlichen Studien eingesetzt werden [Wa65]. Dabei geben Probanden in Abhängigkeit von Münzwürfen oder Ähnlichem zufallsbestimmte oder wahrheitsgemäße Antworten. So lässt sich aus der einzelnen Antwort keine Wahrheit ablesen, d. h. die Privatsphäre der Probanden wird schon bei der Datenerhebung geschützt. Durch das Gesetz der großen Zahlen kann der Einfluss der Zufallsantworten auf die Gesamtheit der Antworten näherungsweise herausgerechnet werden, so dass mit statistischen Methoden Erkenntnisse aus den Daten abgeleitet werden können. Die Anforderungen von Differential Privacy werden bei geeignetem Studienaufbau in der Tat erfüllt [Ka08]. Dadurch ist ein effektiver Schutz der Privatsphäre gegeben und zudem ist das Verfahren rechentechnisch (aber nicht unbedingt für die Probanden) effizient. Zu beachten bleibt aber, dass ein deutlicher Informationsverlust entsteht und dass die Daten in ihren statistischen Eigenschaften hochgradig verändert werden. Letzteres kann rechnerisch korrigiert werden, aber ersteres kann nur mit einer größeren Probandenzahl kompensiert werden.

4 Anonymisierung von Texten

Bei Textdokumenten unterscheiden wir zwischen der Metadatenebene, der Inhaltsebene und der Schreibstilebene. Auf all diesen Ebenen können Personenbezüge vorhanden sein. Bevor wir auf die Anonymisierung hinsichtlich jeder einzelnen Ebene eingehen, klären wir zunächst diese Begriffe. Die *Metadatenebene* ist eine vom Text entkoppelte Ebene, die Zusatzinformationen zu einem Dokument bereitstellt. Die *Inhaltsebene* ist die zentrale Ebene, die die eigentliche Information trägt. Die *Schreibstilebene* ist in die Inhaltsebene eingebettet und lässt sich nicht ohne Weiteres von dieser entkoppeln.

4.1 Anonymisierung auf der Metadatenebene

Die Existenz und Form von Metadaten hängt davon ab, in welchem Format ein Dokument vorliegt. Handelt es sich um eine Datei in einem komplexen Format (z. B. eine PDF-Datei oder ein Word-Dokument), so liegen in der Regel Metadaten vor. Diese enthalten Felder wie etwa Autoren, Titel, Schlüsselwörter und Erstellungsdatum und reichern das Dokument mit semantischen Informationen an. Handelt es sich jedoch um eine reine Textdatei, so existiert innerhalb der Datei keine Metadatenebene. Gegebenenfalls finden sich jedoch Metadaten im umgebenden System, welches die Datei speichert, was beispielsweise ein Dateisystem oder eine E-Mail sein kann.

Metadaten bergen die Gefahr, dass sie oft vom Ersteller nicht wahrgenommen werden, jedoch Informationen enthalten, die dessen Identität ungewollt preisgeben können. Die Anonymisierung der Metadatenebene ist meist trivial durchführbar, indem die Metadaten entweder gar nicht erst erstellt oder nachträglich entfernt werden.

4.2 Anonymisierung auf der Inhaltsebene

Inhaltsdaten enthalten oftmals Entitäten wie z. B. Personennamen, Bezeichnungen von Firmen oder Organisationen oder geographische Orte, die die Identität des Autors oder die von Dritten referenzieren können. Diese lassen sich anders als Metadaten nicht mit einfachen Mitteln entfernen,[2] ohne die Semantik des Dokuments zu verletzen. Die Voraussetzung für die Anonymisierung von Texten ist, zunächst die Verweise auf Identitäten zu identifizieren. Diese können mithilfe computerlinguistischer Verfahren wie *Eigennamenerkennung* (engl. *named entity recognition*) ermittelt werden [Li18a; YB18]. Anschließend können diese Verweise mit verschiedenen Strategien anonymisiert werden. Eine hundertprozentige Erkennung aller Verweise ist jedoch nicht möglich, sodass immer ein Restrisiko verbleibt.

Eine Möglichkeit zur Anonymisierung entsprechender Textstellen läuft über eine Pseudonymisierung mittels partieller Verschlüsselung. Dabei werden Verweise auf Identitäten mit einem geheimen Schlüssels k verschlüsselt, sodass aus dem Dokument $\mathcal{D}$ ein modifiziertes Dokument $\mathcal{D}'$ entsteht. $\mathcal{D}'$ kann somit nur von autorisierten Personen, die k besitzen, entschlüsselt und dadurch vollständig gelesen werden. Stellt man sicher, dass nach der Pseudonymisierung niemand mehr den Schlüssel k hat, ist eine Anonymisierung erreicht. Der Nachteil der partiellen Verschlüsselung ist, dass der Lesefluss in $\mathcal{D}'$ durch die verschlüsselten Elemente gestört wird und das Dokument daher nur fragmentarisch gelesen werden kann, was den Nutzen des Dokuments reduziert.

Eine Alternative zur partiellen Verschlüsselung ist, die Verweise auf Identitäten zu *paraphrasieren*. Damit kann eine Anonymisierung erreicht und gleichzeitig die Semantik von $\mathcal{D}$ bis zu einem gewissen Grad beibehalten werden. Analog zur partiellen Verschlüsselung führt dies zwar ebenfalls zu einem Informationsverlust, allerdings in einer Form, bei der zum einen die modifizierte Version $\mathcal{D}'$ vollständig lesbar bleibt und zum anderen niemand mit einer Art Schlüssel die Ursprungsinformation wiederherstellen kann. Dazu gilt es, die identifizierten Entitäten durch generischere Angaben[3] zu ersetzen.

Eine wichtige Frage bei der Paraphrasierung ist, woher die abgewandelten Entitäten bezogen werden können. Eine Möglichkeit besteht darin, vorhandene linguistische Ressourcen zu verwenden wie etwa *Ontologien* oder *lexikalische Wortnetze*, mit denen semantisch sinnvolle Ersetzungen durchgeführt werden können. Diese müssen in der Regel händisch erstellt werden und sind dadurch mit entsprechenden Aufwand und hohen Kosten verbunden. Hinzu

[2] Ausgenommen sind isolierte Entitäten, die unabhängig vom Text sind (z. B. der Name nach einer Grußformel).
[3] Beispielsweise „Angela Merkel“ → { „deutsche Politikerin“, „gebürtige Hamburgerin“, . . . }.

kommt die Problematik der temporalen Veränderung von Sprachen,[4] sodass gegebenenfalls zu einer Entität x in einem Text keine passenden Ersetzungen in einer Wortliste gefunden werden können, da die Wortliste zu einem Zeitpunkt erstellt wurde, als x noch nicht existierte. Alternativ zu händisch erstellten linguistischen Ressourcen eignen sich Ansätze basierend auf sogenannte *Word Embeddings*. Die Idee dahinter ist, Wörter eines Vokabulars als reelle Vektoren in einem hochdimensionalen Raum darzustellen und diesen auf einen Raum mit niedrigerer Dimension abzubilden, sodass im zweiten Raum semantische Beziehungen der Wörter durch die Nähe der entsprechenden Vektoren widergespiegelt werden. Mithilfe solcher Word Embeddings lassen sich ohne den Einsatz gelabelter Daten bzgl. einer Entität x semantisch ähnliche Entitäten $y_1, y_2, \ldots$ finden, die eine Ersetzung erlauben. Vorausgesetzt werden hier jedoch genügend ungelabelte Textdaten, welche Informationen über die Entität x enthalten. Ein wesentlicher Nachteil hierbei ist allerdings, dass die Entitäten nicht in einer festgelegten Relation (z. B. Synonymie) zueinander stehen, sondern sich über mehrere Relationen wie etwa Hyperonymie, Hyponymie, Meronymie oder Holonymie erstrecken können. Der Literatur zufolge existiert noch kein zufriedenstellender Ansatz mit dessen Hilfe Entitäten hinsichtlich ihrer semantischen Relationen automatisiert abgegrenzt werden können, sodass es hierfür noch weitere Forschungsarbeit bedarf.

4.3 Anonymisierung auf der Schreibstilebene

Die Identität einer Person lässt sich auch über dessen Schreibstil bestimmen. Im Laufe des letzten Jahrzehnts hat sich die *digitale Textforensik* als Forschungsfeld etabliert. Hauptaugenmerk liegt dabei auf der *Autorschaftsanalyse*, welche das Ziel verfolgt, Informationen über die Autoren digitaler Dokumente offenzulegen [Po19].

Aus der Notwendigkeit heraus, die Identität von Autoren zu schützen, entstand das Forschungsfeld *Author Obfuscation* (AO), welches sich damit befasst, wie sich der Schreibstil in Dokumenten verschleiern lässt. Bisherige AO-Ansätze lassen sich in manuelle, computerassistierte und automatische Verfahren aufteilen [GA19], wobei der Forschungsfokus insbesondere auf letzteren liegt. Automatische AO gilt als sehr anspruchsvoll, da sie auf Sprachkompetenzen zurückgreifen muss, um anonymisierende Umformungen in den Dokumenten vorzunehmen unter gleichzeitiger Beibehaltung der ursprünglichen Semantik.

Unter den veröffentlichten automatischen AO-Verfahren ist vor allem der Ansatz *Adversarial Author Attribute Anonymity Neural Translation* (A^4NT) von Shetty et al. [SSF18] hervorzuheben. Das Verfahren ist unserer Recherche nach das einzige Verfahren, das eine dedizierte Komponente für die Semantikerhaltung enthält. A^4NT verfolgt eine intuitive Idee, die analog zu einer maschinellen Übersetzung funktioniert. Während in der maschinellen Übersetzung ein Dokument in eine festgelegte Zielsprache übersetzt wird, wird bei A^4NT das Dokument in dieselbe Sprache wie die Quellsprache „übersetzt“, um den Schreibstil des ursprünglichen Autors nicht mehr wiedererkennen zu können. Das Verfahren wurde

[4] Vor 20 Jahren gab es z. B. noch nicht die Wörter „googlen“, „Podcast“ und „Smombie“.

hinsichtlich der drei autorspezifischen Attribute Alter (unter 20 vs. über 20), Geschlecht und Identität (Obama vs. Trump) anhand einer Kollektion von Blogartikeln und einer Kollektion von politischen Reden getestet. Hinsichtlich der Attribute Alter und Geschlecht konnten Shetty et al. die Erkennungsgenauigkeit (F_1-Wert) beim Alter von 88 % auf 8 %, beim Geschlecht von 75 % auf 39 % und bei der Identität von 100 % auf 0 % senken, was dafür spricht, dass eine Anonymisierung auf der Schreibstilebene möglich ist.

5 Anonymitätsrisiken beim maschinellen Lernen

Im Zeitalter von *Big Data* und *maschinellem Lernen* (ML) ist es noch schwieriger geworden, Privatheit zu gewährleisten, da in großen Datenbeständen – selbst in solchen aus gering strukturierten oder gar unstrukturierten Daten – die entscheidenden Verknüpfungen gefunden werden können, welche das Herstellen von Personenbezügen ermöglichen. Da ML-Algorithmen üblicherweise auf disjunkten Datensätzen trainiert und evaluiert werden, wurde lange fälschlicherweise angenommen, dass es nicht möglich ist, vom finalen Modell Rückschlüsse auf die zum Training verwendeten Daten zu ziehen. Bestimmte ML-Techniken können sich jedoch unerwartet deutlich an die zum Training des Modells verwendeten Daten erinnern. So speichern Support Vector Machines oder k-nächste-Nachbarn-Klassifikatoren Informationen über die zum Lernen verwendeten Daten in dem Modell selbst ab. Diese sogenannten Feature-Vektoren erlauben unter bestimmten Umständen Rückschlüsse auf die Rohdaten und stellen somit ein entscheidendes Risiko dar [AC19].

Fredrikson et al. [FJR15] demonstrierten, dass die Erinnerung in neuronalen Netzen, welche zur Gesichtserkennung genutzt wurden, mitunter so stark sein kann, dass es möglich ist, ein Abbild der Trainingsdaten zu rekonstruieren – ein sogenannter *Modellinversionsangriff*. Shokri et al. [Sh17] bewiesen, dass neuronale Netze aufgrund ihrer Konstruktion anfällig für *Membership-Inference-Angriffe* sind. Die Autoren wiesen nach, dass ein trainiertes Netz merkbar anders auf Informationen reagiert, welche bereits zum Training verwendet wurden als auf bisher ungesehene Testdaten. Aufgrund dieser Rückmeldung kann ein Angreifer eindeutig zuordnen, ob ein Individuum in einem bestimmten Datensatz enthalten ist oder nicht. Solche Angriffe stellen allgemein eine Verletzung der Privatheit dar, sind aber besonders dann kritisch, wenn es sich um sensible Informationen handelt, wie beispielsweise Insolvenz oder ob eine bestimmte Krankheit vorliegt.

Das Forschungsfeld *Privacy Preserving Machine Learning* (PPML) ist noch recht jung. Auch wenn es bereits vielversprechende Ansätze gibt, besteht noch viel Entwicklungsbedarf. Nachfolgend werden die wichtigsten Forschungsrichtungen auf diesem Gebiet skizziert.

5.1 Kollaboratives maschinelles Lernen

Würden alle oder viele Personen ihre Daten nicht mehr für Forschungs- und Auswertungszwecke zur Verfügung stellen, hätte das einschneidende Konsequenzen für die Forschung,

insbesondere für die Medizinforschung. Außerdem könnten viele weitverbreitete, nützliche Dienste nicht weiter angeboten werden. Ziel ist es folglich, Daten auf privatsphärenfreundliche Weise einem ML-System zur Verfügung stellen zu können.

Kryptographische Verfahren für kollaboratives Lernen Ein vielversprechender Ansatz, um die Privatheit des Einzelnen zu schützen und gleichzeitig das Training von Modellen auf Daten von vielen Personen zu ermöglichen, ist die *sichere Mehrparteienberechnung* (engl. *secure multi-party computation*, MPC) als Teilgebiet der Kryptographie. Das Ziel von MPC ist das gemeinschaftliche Berechnen einer Funktion, für die mehrere Parteien eine Eingabe liefern. Die Privatheit wird in dieser Art der Berechnung dadurch gewahrt, dass jede der beteiligten Parteien nur das Endergebnis, d. h. die Funktionsausgabe, und die eigene Eingabe erfährt. Die Eingaben der übrigen Teilnehmer bleiben verborgen. Je nach Anzahl der Teilnehmer und deren Abbruchwahrscheinlichkeit existieren verschiedene Ansätze mit unterschiedlichem Rechen- und Kommunikationsaufwand, um dieses Ziel zu erreichen. Tatsächlich gibt es erste MPC-Ansätze im Kontext von maschinellem Lernen zur Summenberechnung von Modellparametern [Bo17].

Homomorphe Verschlüsselung erlaubt – im Gegensatz zu herkömmlichen Verschlüsselungsmethoden – Rechenoperationen direkt auf den verschlüsselten Daten auszuführen, ohne diese zuvor in Klartext zu überführen und sie dadurch angreifbar zu machen. Jede Operation liefert ein ebenfalls verschlüsseltes Ergebnis, das dechiffriert demjenigen entspricht, welches resultieren würde, wäre die Operation auf dem entsprechenden Klartext durchgeführt worden. Mit homomorpher Verschlüsselung können daher Daten an eine nicht-vertrauenswürdige Instanz weitergegeben werden und Berechnungen dort durchgeführt werden. Insbesondere die sogenannte voll-homomorphe Verschlüsselung generiert jedoch einen signifikanten Rechenmehraufwand [Do16; Li18b], der diese für rechenintensive Anwendungen wie maschinelles Lernen bisher unbrauchbar macht. Erste praktikable Ansätze verwenden daher Vereinfachungen. So wenden Dowlin et al. [Do16] ein auf unverschlüsselten Rohdaten trainiertes neuronales Netz auf „somewhat“ homomorph verschlüsselte Daten an. Long et al. [Lo18] haben das Training verschiedener ML-Verfahren mit additiv-homomorpher Verschlüsselung und Zero-Knowledge-Beweisen realisiert.

Dezentrales maschinelles Lernen Eine Lösung zum privatsphärenfreundlichen Lernen auf Daten von vielen Nutzern ist das dezentrale Lernen. Hierbei trainieren die Nutzer ein Grundmodell lokal auf ihren individuellen Daten und übermitteln lediglich die neu berechneten Gradienten des Trainings oder die neuen Modellparameter an den Serviceprovider. In einem periodischen Prozess aktualisiert der Provider das Gesamtmodell anhand der übermittelten Informationen aller Teilnehmer und stellt es ihnen anschließend zum Download zur Verfügung. Diese trainieren nun das aktualisierte Modell erneut lokal und senden die resultierenden Gradienten oder Parameter zurück an den Server [Mc17]. Hauptsächlich zum Schutz der Privatsphäre, aber auch zur Kommunikationseffizienz, erlaubt der

Ansatz nach Shokri und Shmatikov [SS15], dass nicht alle Aktualisierungen mit dem Server geteilt werden müssen, sondern nur eine kleine Teilmenge, deren Größe vom Nutzer selbst festgelegt wird. Allerdings sollte sich der Nutzer des Trade-offs zwischen der Menge der geteilten Aktualisierungen sowie Trainingszeit und -qualität bewusst sein.

Hitaj et al. [HAP17] haben nachgewiesen, dass es selbst in solchen dezentralen Lernansätzen mit Hilfe eines Generative Adversarial Networks (GAN) möglich ist, über die übrigen aufrichtigen Teilnehmer sensible Daten zu sammeln. Melis et al. [Me19] entkräften teilweise die Angriffe von Hitaj et al., zeigen aber selbst neue Angriffsstrategien auf.

5.2 Differential Privacy für maschinelles Lernen

Arbeiten zu *Differentially Privacy* im Kontext des maschinellen Lernens (vgl. Differential Privacy in Abschnitt 3) erforschen verschiedene Aspekte des Verrauschens von potentiell angreifbaren Daten. Untersucht wird hier meist, auf welcher Ebene die Störungen idealerweise Eingang in den Algorithmus finden – ob nun auf Input- oder Output-Ebene oder ob die Gradienten oder die Verlustfunktion verrauscht werden – und welche Verteilungseigenschaften das Rauschen selbst haben sollte. Das Ziel ist, einen optimalen Trade-off zwischen Privatheit und Ergebnisqualität zu erreichen.

Eine andere Richtung verfolgt der Ansatz der *Differentially Private Data Synthesis* (DIPS). Hierbei werden Daten auf Basis realer Datensätze beispielsweise mittels Copula-Funktionen [LXJ14] oder Generative Adversarial Networks [TF19] unter Einhaltung von Differential Privacy synthetisiert. Der offensichtliche Vorteil dieses Ansatzes ist, dass die simulierten Daten bereits Differential Privacy erfüllen und somit keine Rückschlüsse auf die Ursprungsdaten ermöglichen – im Gegensatz zu anderen Datensyntheseverfahren. Darüber hinaus besitzen die Daten annähernd die gleichen Verteilungseigenschaften wie die zugrundeliegenden Originaldaten und können in beliebiger Anzahl generiert werden, um so beispielsweise die Güte eines ML-Models zu verbessern [ML19; PCN18].

5.3 Generelle Limitationen der bestehenden Verfahren

Angriffspunkt für die weitere Forschung ist u. a. die Anwendbarkeit der Verfahren bzw. deren mangelnde Flexibilität. Die meisten privatheiterhaltenden Verfahren sind nur für die Anwendung auf einen bestimmten Lernalgorithmus optimiert und auf andere ML-Verfahren schwer bis gar nicht übertragbar. Zudem stellt mangelnde Skalierbarkeit ein Hindernis für die Anwendung privatheiterhaltender Maßnahmen in der Praxis dar. Das Schützen sensibler Informationen generiert immer zusätzliche Kosten – entweder aufgrund von höherem Berechnungsaufwand, extrem langen Trainingszeiten oder weil der Nutzen der Daten bspw. durch zugefügtes Rauschen vermindert wird. In manchen Fällen fallen diese Kosten sogar so groß aus, dass eine Anwendung in der Praxis nicht tragbar ist [AC19].

6 Zusammenfassung der Herausforderungen und Fazit

Grundsätzlich ist eine exakte Definition von Personenbezug und Anonymität nötig, an der die rechtliche Einordnung von Daten zweifelsfrei entschieden werden kann und an der Anonymitätskriterien zur praktischen Prüfung von Daten gemessen werden können. Zudem ist eine Weiterentwicklung auf dem Gebiet der Anonymitätskriterien nötig. Zum einen existieren solche Kriterien hauptsächlich für tabellarische Daten. Zum anderen mangelt es den meisten dieser Kriterien an starken Garantien, so dass etwa Angreifer mit zusätzlichem Hintergrundwissen weitere Informationen über konkrete Personen extrahieren können. Daher müssen diese Kriterien durch eine theoretische Untersuchung basierend auf zu entwickelnden formalen Angreifermodellen hinsichtlich ihrer Garantien präzisiert werden.

Für strukturierte Daten sind die heutigen Kernmethoden zur Anonymisierung hauptsächlich vor zehn bis zwanzig Jahren publiziert worden und es wurden bereits viele Verbesserungen entwickelt. Handlungsbedarf besteht aber weiterhin neben den bereits genannten Problemen der Anonymitätskriterien auch in Bezug auf die Minimierung des Informationsverlustes bei gleichzeitiger Maximierung der Effizienz von Algorithmen, insbesondere in Bezug auf eine adaptive Ausbalancierung dieser entgegengesetzten Anforderungen.

Bei der Anonymisierung der Inhaltsebene von Textdaten gibt es nach wie vor die Herausforderung der zuverlässigen Erkennung von Entitäten sowie Herausforderungen in Bezug auf Umsetzbarkeit und Anwendbarkeit von Strategien zur Ersetzung dieser Entitäten. In Bezug auf die Anonymisierung der Stilebene gibt es erste empirische Ergebnisse, die erfolgversprechend sind, aber eine allgemeine Zuverlässigkeit kann noch nicht daraus geschlossen werden. Zudem ist eine Anonymitätsgarantie bei Texten noch weniger möglich als bei strukturierten Daten.

Der Privatsphärenschutz in Verbindung mit maschinellem Lernen ist ein noch recht junges und unerforschtes Thema, das erst vor wenigen Jahren verschiedene Risiken aufgedeckt hat. Erste Lösungsansätze, etwa in Verbindung mit Differential Privacy oder Kryptographie, beschränken sich hauptsächlich auf den Schutz additiver Operationen. Neben den allgemeinen Herausforderungen dieser Schutzstrategien sind hier auch die Herausforderungen der Anwendbarkeit und Effektivität im Kontext des maschinellen Lernens zu lösen.

Diese Arbeit hat gezeigt, dass es bereits viele wissenschaftliche Publikation zur Anonymisierung von Daten gibt. Die Literatur behandelt sowohl Anonymisierungskonzepte als auch Limitationen und Schutzlücken der Konzepte. Bei strukturierten Daten ist unter Berücksichtigung der Einschränkungen in Bezug auf Anonymitätsgarantien und Algorithmeneffizienz ein Praxiseinsatz von Anonymisierung bereits möglich, während auf den anderen untersuchten Gebieten die Anonymisierungsstrategien hauptsächlich prototypische Forschungsarbeiten sind. In allen Bereichen gibt es noch viele zu lösende Forschungsfragen, die hier herausgearbeitet wurden.

Literatur

[AC19] Al-Rubaie, M.; Chang, J. M.: Privacy-Preserving Machine Learning: Threats and Solutions. IEEE Security & Privacy 17/2, S. 49–58, 2019.

[BLR08] Blum, A.; Ligett, K.; Roth, A.: A Learning Theory Approach to Non-Interactive Database Privacy. In: STOC'08. ACM, S. 609–618, 2008.

[Bo17] Bonawitz, K. et al.: Practical Secure Aggregation for Privacy-Preserving Machine Learning. In: CCS 2017. ACM, S. 1175–1191, 2017.

[Do16] Dowlin, N. et al.: CryptoNets: Applying Neural Networks to Encrypted Data with High Throughput and Accuracy, Techn. Ber., 2016.

[DSG14] Artikel-29-Datenschutzgruppe: Opinion 05/2014 on Anonymisation Techniques, Techn. Ber. WP216, Artikel-29-Datenschutzgruppe, 10. Apr. 2014.

[Du07] Du, Y. et al.: On Multidimensional k-Anonymity with Local Recoding Generalization. In: ICDE 2007. IEEE, 2007.

[Dw06a] Dwork, C.: Differential Privacy. In: Automata, Languages and Programming. Springer, S. 1–12, 2006.

[Dw06b] Dwork, C. et al.: Calibrating Noise to Sensitivity in Private Data Analysis. In: Theory of Cryptography. Springer, S. 265–284, 2006.

[FJR15] Fredrikson, M.; Jha, S.; Ristenpart, T.: Model inversion attacks that exploit confidence information and basic countermeasures. In: CCS 2015. ACM, S. 1322–1333, 2015.

[GA19] Gröndahl, T.; Asokan, N.: Text Analysis in Adversarial Settings: Does Deception Leave a Stylistic Trace?, 26. Feb. 2019, arXiv: 1902.08939v2.

[HAP17] Hitaj, B.; Ateniese, G.; Pérez-Cruz, F.: Deep models under the GAN: information leakage from collaborative deep learning. In: CCS 2017. ACM, S. 603–618, 2017.

[Ka08] Kasiviswanathan, S. P. et al.: What Can We Learn Privately? In: FOCS 2008. IEEE Computer Society, S. 531–540, 2008.

[Li18a] Li, J. et al.: A Survey on Deep Learning for Named Entity Recognition, 22. Dez. 2018, arXiv: 1812.09449v1.

[Li18b] Liu, Q. et al.: A survey on security threats and defensive techniques of machine learning: A data driven view. IEEE Access 6/, S. 12103–12117, 2018.

[LLV07] Li, N.; Li, T.; Venkatasubramanian, S.: t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In: ICDE 2007. IEEE, S. 106–115, 2007.

[Lo18] Long, Y. et al.: Distributed and Secure ML with Self-tallying Multi-party Aggregation, 26. Nov. 2018, arXiv: 1811.10296v1.

[LXJ14] Li, H.; Xiong, L.; Jiang, X.: Differentially private synthesization of multi-dimensional data using copula functions. In: EDBT 2014. OpenProceedings, S. 475–486, 2014.

[Ma06] Machanavajjhala, A. et al.: *l*-Diversity: Privacy Beyond *k*-Anonymity. In: ICDE'06. IEEE, Apr. 2006.

[Mc17] McMahan, H. B. et al.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: AISTATS 2017. Bd. 54. PMLR, S. 1273–1282, 2017.

[Me19] Melis, L. et al.: Exploiting unintended feature leakage in collaborative learning. In: IEEE S&P 2019. 2019.

[ML19] McKay Bowen, C.; Liu, F.: Comparative study of differentially private data synthesis methods, 8. Jan. 2019, arXiv: 1602.01063v4.

[MT07] McSherry, F.; Talwar, K.: Mechanism Design via Differential Privacy. In: FOCS 2007. IEEE Computer Society, S. 94–103, 2007.

[NAC07] Nergiz, M. E.; Atzori, M.; Clifton, C. W.: Hiding the Presence of Individuals from Shared Databases. In: SIGMOD'07. ACM, S. 665–676, 2007.

[PCN18] Page, H.; Cabot, C.; Nissim, K.: Differential privacy an introduction for statistical agencies. In: NSQR. Government Statistical Service, 2018.

[Po19] Potthast, M. et al.: A Decade of Shared Tasks in Digital Text Forensics at PAN. In: Advances in Information Retrieval. Springer, S. 291–300, 2019.

[Sh17] Shokri, R. et al.: Membership inference attacks against machine learning models. In: IEEE S&P 2017. S. 3–18, 2017.

[SS15] Shokri, R.; Shmatikov, V.: Privacy-preserving deep learning. In: ACM CCS 2015. ACM, S. 1310–1321, 2015.

[SSF18] Shetty, R.; Schiele, B.; Fritz, M.: A^{4}NT: Author Attribute Anonymity by Adversarial Training of Neural Machine Translation. In: USENIX Security '18. S. 1633–1650, 2018.

[Sw01] Sweeney, L.: Computational Disclosure Control, A Primer on Data Privacy Protection, Diss., Massachusetts Institute of Technology, Mai 2001.

[TF19] Triastcyn, A.; Faltings, B.: Generating artificial data for private deep learning. In: PAL. Bd. Vol-2335. CEUR Workshop Proceedings, S. 33–40, 18. März 2019.

[Wa65] Warner, S. L.: Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. JASA 60/309, S. 63–69, 1965.

[YB18] Yadav, V.; Bethard, S.: A Survey on Recent Advances in Named Entity Recognition from Deep Learning models. In: COLING 2018. ACL, S. 2145–2158, Aug. 2018.

Konzepte zum Schutz privater Muster in Zeitreihendaten

IoT-Anwendungen im Spannungsfeld zwischen Servicequalität und Datenschutz

Christoph Stach[1]

Abstract: Obwohl das *Internet der Dinge* (*IoT*) die Voraussetzung für *smarte* Anwendungen schafft, die signifikante Vorteile gegenüber traditionellen Anwendungen bieten, stellt die zunehmende Verbreitung von IoT-fähigen Geräten auch eine immense Gefährdung der Privatheit dar. IoT-Anwendungen sammeln eine Vielzahl an Daten und senden diese zur Verarbeitung an ein leistungsstarkes Back-End. Hierbei werden umfangreiche Erkenntnisse über den Nutzer gewonnen. Erst dieses Wissen ermöglicht die Servicevielfalt, die IoT-Anwendungen bieten. Der Nutzer muss daher einen Kompromiss aus Servicequalität und Datenschutz treffen. Heutige Datenschutzansätze berücksichtigen dies unzureichend und sind dadurch häufig zu restriktiv. Aus diesem Grund stellen wir neue Konzepte zum Schutz privater Daten für das IoT vor. Diese berücksichtigen die speziellen Eigenschaften der im IoT zum Einsatz kommenden *Zeitreihendaten*. So kann die Privatheit des Nutzers gewährleistet werden, ohne die Servicequalität unnötig einzuschränken. Basierend auf den *TICK-Stack* beschreiben wir Implementierungsansätze für unsere Konzepte, die einem *Privacy-by-Design-Ansatz* folgen.

Keywords: Datenschutz; Zeitreihendaten; IoT; DSGVO; ePrivacy-Verordnung; TICK-Stack.

1 Einleitung

Das *Internet der Dinge* (engl. *Internet of Things*, kurz *IoT*) war lange Zeit nicht viel mehr als ein Modewort und Wissenschaft und Wirtschaft versuchten, das damit verbundene Potential anhand von kleineren Pilotprojekten auszuloten. Hierbei zeigte sich, dass die im IoT erfassten Sensordaten in vielen Anwendungsbereichen gewinnbringend eingesetzt werden können, wie beispielsweise im *Smart Home*, im *Gesundheitssektor* oder zur Realisierung der *Industrie 4.0*. Gartner Analysten sehen das Jahr 2017 jedoch als Gipfel des *Hype-Zyklus* für das IoT an und in den kommenden Jahren gilt es, die erarbeiteten *Proof-of-Concept-Projekte* mithilfe von marktfähigen Umsetzungen zu untermauern [Ve17]. Während die Verfügbarkeit von IoT-fähigen Geräten sowohl im privaten als auch im industriellen Umfeld bereits weitestgehend gegeben ist, gibt es softwareseitig noch viele offene Fragen.

Insbesondere die Frage, wie relevantes Wissen aus den Unmengen an verfügbaren Daten gewonnen werden kann, ist entscheidend für die Servicequalität einer IoT-Anwendung. Da IoT-fähige Geräte in der Regel nicht ausreichend Ressourcen besitzen, um selbstständig umfassende Analysen auszuführen, senden sie die erfassten Daten hierzu an eine zentrale

[1] Universität Stuttgart, Universitätsstraße 38, 70569 Stuttgart, Deutschland, Christoph.Stach@ipvs.uni-stuttgart.de

doi:10.18420/inf2019_54

Verarbeitungskomponente. Hier werden alle erfassten Daten miteinander verknüpft, bereinigt und mittels *Data-Mining-* respektive *Machine-Learning-Techniken* analysiert. Damit die Verarbeitung der Daten echtzeitnah erfolgen kann, muss das Datenvolumen frühzeitig reduziert werden. Hierfür lassen sich die speziellen Eigenschaften dieser *Zeitreihendaten* ausnutzen. Anstelle sämtliche Datenpunkte zu analysieren, genügt es beispielsweise häufig, in den Daten enthaltene charakteristische Muster zu betrachten (z. B. Extremwerte) [Ma17].

Eine noch größere Herausforderung stellt jedoch der Schutz privater Daten dar. Einerseits leben innovative IoT-Anwendungen davon, uneingeschränkten Zugriff auf Nutzerdaten zu erhalten und auf diese Weise ihre Services auf die Bedürfnisse des Nutzers[2] auszurichten. Andererseits stellt just dieser uneingeschränkte Zugriff in Kombination mit den nahezu grenzenlosen Ressourcen, die zur Analyse der Daten zur Verfügung stehen, eine ernstzunehmende Gefährdung der Privatsphäre dar. Während aus rechtlicher Sicht beispielsweise die *DSGVO* den Zugriff und die Verarbeitung von personenbezogenen Daten regelt, bedarf es technischer Lösungen zur Sicherstellung der Datenschutzbestimmungen. Es muss dem Nutzer möglich sein, einfach festzulegen, welche seiner Daten wofür verwendet werden dürfen und ebenfalls, welche Informationen nicht preisgegeben werden dürfen [Zh18].

Hierbei zeigt sich das Spannungsfeld, dem sich IoT-Anwendungen ausgesetzt sehen: Ein Nutzer muss zwischen Servicequalität und Datenschutz abwägen. Je mehr Daten er einer IoT-Anwendung zur Analyse bereitstellt, desto mehr Funktionalität kann sie ihm anbieten – gleichzeitig werden allerdings auch viele private Informationen offengelegt. Erhält eine IoT-Anwendung hingegen keine Daten, ist die Privatheit des Nutzers geschützt, aber die Anwendung wird ineffektiv. Ziel muss es daher sein, einen guten Kompromiss zwischen Servicequalität und Datenschutz zu finden. Hierbei kann davon profitiert werden, dass die erfassten Daten im Rahmen der Analyse zur Reduktion des Datenvolumens ohnehin verdichtet werden müssen. Durch eine geschickte Wahl des Verdichtungsverfahrens, können private Informationen verborgen werden, ohne die Analyseergebnisse zu kompromittieren.

Zu diesem Zweck erbringen wir die folgenden drei Beiträge: (1) Wir stellen sechs Konzepte zum Schutz privater Muster in Zeitreihendaten vor, die die speziellen Eigenschaften dieser Daten ausnutzen. Mithilfe unserer Konzepte ist es möglich, die Privatheit des Nutzers zu gewährleisten und gleichzeitig die Servicequalität von IoT-Anwendungen nicht unnötig einzuschränken. (2) Wir bewerten die eingeführten Konzepte und schätzen ab, welches Konzept sich für welchen Anwendungsfall sich am besten eignet. (3) Wir beschreiben auf den *TICK-Stack*[3] basierende Implementierungsansätze für unsere Schutzkonzepte.

Der Rest dieses Artikels ist wie folgt gegliedert: In Abschnitt 2 wird der Stand der Technik bezüglich IoT-Anwendungen anhand eines Smart-Home-Anwendungsfalls beschrieben. Datenschutzansätze fürs IoT werden in Abschnitt 3 diskutiert. In Abschnitt 4 führen wir neue Datenschutzkonzepte ein und evaluieren diese. Wir diskutieren Implementierungsansätze für diese Schutzkonzepte in Abschnitt 5. Abschließend fasst Abschnitt 6 den Artikel zusammen.

[2] Mit dem Begriff „Nutzer“ seien im Folgenden jeweils alle Geschlechter gleichermaßen adressiert.
[3] siehe https://www.influxdata.com/time-series-platform/

2 Stand der Technik

Im Nachfolgenden führen wir ein Smart-Home-Anwendungsbeispiel ein. Anhand dieses Beispiels beschreiben wir anschließend die technischen Grundlagen, die eine solche IoT-Anwendung ermöglichen, und gehen darauf ein, welche rechtlichen Rahmenbedingungen die DSGVO sowie die *ePrivacy-Verordnung* für solche Anwendungen schaffen.

Anwendungsbeispiel: In Anlehnung an Marikyan et al. [Ma19] eruierten wir folgende drei Nutzungsmöglichkeiten für IoT-Anwendungen in einem Smart Home. So können IoT-fähige Geräte ihre Nutzer bei der *Bewältigung alltäglicher Aufgaben* unterstützen. Hierzu werden mittels Sensoren die gegenwärtigen Aktivitäten der Nutzer erfasst und Aktuatoren reagieren darauf. Denkbare Anwendungsszenarien hierfür sind, dass Senioren automatisch darauf hingewiesen werden können, wenn die Aktivität „Einnahme von Medikamenten“ ausblieb oder dass die Heizung automatisch deaktiviert wird, wenn ein Nutzer ein Fenster öffnet.

IoT-Technologien lassen sich auch dafür nutzen, die *empfundene Lebensqualität* der Nutzer zu steigern. Kameras können besondere Ereignisse, wie beispielsweise eine Party, in Bild und Ton festhalten. Nutzer können diese Aufnahmen indexieren. So ist das Smart Home in der Lage, darauf abgebildete Personen zu erkennen oder neue Aufnahmen automatisch bestimmten Rubriken zuzuordnen. Auf digitalen Fotorahmen können situationsabhängig passende Bilder anzeigen (z. B. abhängig von den anwesenden Personen).

Schließlich stellt auch die *Überwachung* des Smart Homes eine Nutzungsmöglichkeit dar. Hat eine IoT-Anwendung alle typischen Aktivitäten der ihr bekannten Nutzer erfasst, kann sie ebenfalls Abweichungen von diesen Verhaltensmustern erkennen. Liegt ein älterer Nutzer beispielsweise an einem ungewöhnlichen Ort, könnte er gestürzt sein und eine Notsituation vorliegen, oder verschafft sich ein unbekannter Nutzer Zugang zum Haus, so könnte es sich um einen Einbruch handeln. In beiden Fällen, kann automatisch Hilfe gerufen werden.

Es ist offensichtlich, dass solche Anwendungen trotz des großen Nutzungspotentials Bedenken bezüglich des Datenschutzes aufwerfen. All diese Services sind nur möglich, wenn die eingesetzten IoT-fähigen Geräte dauerhaft Daten erfassen, austauschen und verarbeiten können. Da aufgrund der großen Datenmenge die verfügbaren Ressourcen weder für die Speicherung noch die Verarbeitung ausreichen, erfolgt dies häufig auf externen Servern des Serviceanbieters. Nutzer müssen sich daher entscheiden, welche Daten sie bereit sind preiszugeben, um im Gegenzug welche Serviceleistungen zu erhalten [Zh18].

Technische Grundlagen: Um eine solche IoT-Anwendung zu ermöglichen, bedarf es aus technischer Sicht zwei wesentliche Komponenten. Zum einen muss es möglich sein, *große Datenmengen* zu verarbeiten (z. B. um historische Daten zu analysieren und darin enthaltene Verhaltensmuster zu lernen) und zum anderen müssen *Daten in Echtzeit* verarbeitet werden können (z. B. um unmittelbar auf festgestellte Muster in den Live-Daten zu reagieren).

Für derartige Anwendungsfälle hat sich die *Lambda-Architektur* [MW15] als De-Facto-Standard durchgesetzt. Diese ist in Abb. 1 dargestellt. Beliebige Datenquellen, wie *Sensoren*

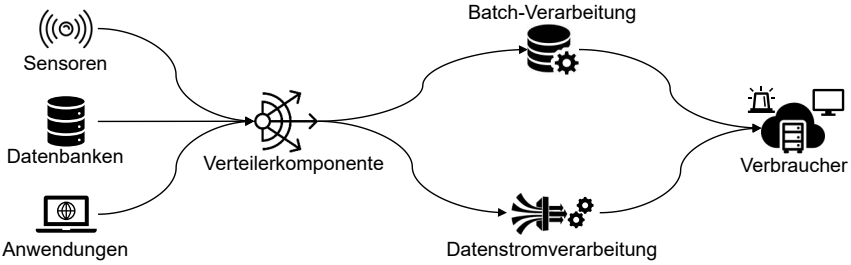


Abb. 1: Die *Lambda-Architektur* in Anlehnung an Marz und Warren [MW15].

(respektive IoT-fähige Geräte), *Datenbanken* oder *Anwendungen* senden ihre Daten an eine *Verteilerkomponente*. Diese reichert eingehende Daten um einen Zeitstempel sowie ein *Topic* an, das den Kontext beschreibt, in dem die Daten erfasst wurden. Aufgrund des Zeitstempels sind sämtliche Daten ab diesem Punkt Zeitreihendaten. Für die Langzeitarchivierung und Historisierung, werden die Daten in einer Datenbank abgelegt. In dieser werden die Daten en gros als *Batch* verarbeitet, wodurch eine hohe Genauigkeit der Analyseergebnisse erzielt wird. Allerdings dauert die Datenverarbeitung dadurch auch mehrere Stunden. Beispielsweise können auf diese Weise Muster gelernt werden, die beschreiben, wie eine bestimmte Aktivität erkannt werden kann. Für die eigentliche Mustererkennung ist dieses Vorgehen aufgrund der hohen Laufzeit jedoch ungeeignet. Daher werden eingehende Daten parallel dazu auch an eine *Datenstromverarbeitungskomponente* geschickt. Hier werden die gelernten Muster auf neue Daten innerhalb eines beschränkten Zeitfensters angewandt. Aufgrund der geringen Datenmenge, die dabei zu jedem Zeitpunkt berücksichtigt wird, sind zwar keine umfassenden Analysen möglich, die Berechnung kann jedoch sehr schnell erfolgen (*echtzeitnah*). Beide Verarbeitungskomponenten stellen ihre Ergebnisse ausgewählten *Verbrauchern* (z. B. Aktuatoren) zur Verfügung. Eine Implementierungsvariante der Lambda-Architektur ist mit dem TICK-Stack in Abschnitt 5 beschrieben.

Rechtliche Situation: Es ist offensichtlich, dass IoT-Anwendungen davon profitieren, wenn sie möglichst viele Daten über den Nutzer erfassen und verarbeiten können. Wachter [Wa18] untersucht daher, inwiefern sich derartige Anwendungen mit der DSGVO vereinbaren lassen. Es ist offensichtlich, dass sich Artikel 5 hierbei als problematisch erweist, da sich eine *Datenminimierung* hochgradig negativ auf die Servicequalität auswirkt. Auch werden Sensordaten im IoT dauerhaft gesammelt, während sich neue Verwendungszwecke erst mit der Zeit ergeben (*Zweckbindung*). Überhaupt ist es quasi unmöglich eine *Transparenz* zu schaffen, da aus den Daten Modelle gelernt werden, die für Menschen nicht nachvollziehbar sind. Artikel 22 reguliert *automatisierte Entscheidungen* und *Profiling*. Da IoT-Anwendungen ohne diese beiden Techniken nicht auskommen, ist eine ausdrückliche *Einwilligung* erforderlich. Die Einwilligungsanfrage muss in *klarer und einfacher Sprache* erfolgen (Artikel 7), was sich aufgrund der komplexen Verarbeitungsweise als zusätzlich schwierig erweist. Die ePrivacy-Verordnung weitet diese Schutzanforderungen außerdem auf *nicht-personenbezogene Daten* sowie *Daten von juristischen Personen* aus [Vo17].

3 Verwandte Arbeiten

Wie der Blick auf den Stand der Technik zeigt, machen die DSGVO und die ePrivacy-Verordnung grundlegende Änderungen an IoT-Anwendungen erforderlich, insbesondere da es für die Anbieter dieser Anwendungen nahezu unmöglich ist, nachträglich nachzuvollziehen, aufgrund welcher Daten eine Aktion ausgelöst wurde. Eine Lösung kann Artikel 25 der DSGVO darstellen. Hier werden explizit *technische Maßnahmen* gefordert, die die Einhaltung des Datenschutzes gewährleisten, d. h. eine IoT-Anwendung soll sich selbst überwachen und regulieren. Im Folgenden untersuchen wir daher, welche Datenschutztechniken aktuell im IoT-Kontext eingesetzt werden und bewerten diese in Hinblick auf den obigen Anwendungsfall.

Zugriffskontrollverfahren: Bourgeois et al. [Bo18] schlagen daher den Einsatz eines rollenbasierten Zugriffskontrollsystems vor. Hierbei werden den an der IoT-Anwendung beteiligten Parteien bestimmte *Rollen* zugeordnet (z. B. betroffene Person, Datenverantwortlicher oder Anfragender). Eine Partei kann auch mehrere Rollen innehaben und diese je nach Anwendungsfall dynamisch wechseln. Jeder Rolle sind Zugriffsrechte auf die Daten(quellen) einer IoT-Anwendung zugeteilt. Reicht diese starre Rollenstruktur nicht aus, können die Zugriffsberechtigungen auch an *Attribute* gebunden werden, die den Nutzer, die angefragten Daten oder den Kontext, in dem ein Nutzer eine Anfrage stellt, beschreiben [Ou16].

Während sich dieser Ansatz in Mehrbenutzersystemen bewährt hat, ergeben sich fürs IoT zwei entscheidende Nachteile: Zum einen kann es dabei leicht zu einer Identitätsfälschung kommen, da die Echtheit der übermittelten Attribute nicht überprüft werden kann [Bh19]. Hierfür wäre eine Verifikation der IoT-fähigen Geräte nötig, die die verwendeten Attribute beisteuern. Dies ist allerdings ressourcenintensiv und kann nicht auf den IoT-fähigen Geräte ausgeführt werden. Gritti et al. [Gr19] stellen daher ein Verfahren vor, mit dem diese Verifikation auf einen *Cloud-Dienst* ausgelagert werden kann. Hierdurch ergibt sich jedoch sowohl für den Anbieter als auch den Nutzer ein zusätzlicher Aufwand. Zum anderen setzt dieser Ansatz voraus, dass sämtliche (private) Daten an den IoT-Anbieter übergeben werden, d. h. sie verlassen den Einflussbereich des Nutzers. Ob und wie gut dort die Zugriffskontrollverfahren angewandt werden, ist für den Nutzer nicht ersichtlich [Bh19].

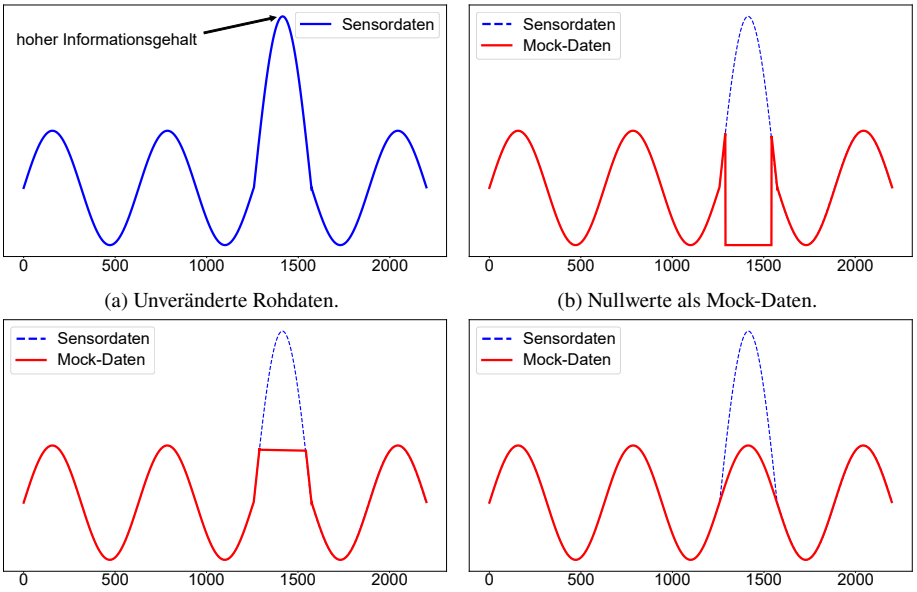
Attributbasierte Verfahren: Um dieser Problematik entgegenzuwirken, kann in der Verteilerkomponente oder der Datenquelle ein *Filter* integriert werden. So können die Daten eines ausgewählten Sensors aus dem Datenstrom herausgefiltert werden, wenn diese private Informationen beinhalten. Ein Nutzer könnte beispielsweise festlegen, dass die Daten eines GPS-Sensors nicht an eine IoT-Anwendung weitergegeben werden. Auch eine feingranulare Filterung ist möglich, indem nur bestimmte Attribute eines Sensors blockiert werden (z. B. der Längen- oder Breitengrad). Der Filter kann zusätzlich mit einem *Kontext* verknüpft werden, der beschreibt, wann der Filter aktiv sein soll (z. B. „nur am Wochenende“). Olejnik et al. [Ol17] stellen ein solches System für IoT-fähige Geräte vor, das ebenfalls für die Verteilerkomponente implementiert werden kann. Selbst wenn die Daten bereits an die Batch- oder Datenstromverarbeitung weitergereicht wurden, gibt es vergleichbare Ansätze, zur Filterung von Datenbankanfragen [PO12] und Datenströmen [Ad11].

Ein Problem, das jedoch all diesen Verfahren inhärent innewohnt, ist deren hohe Restriktivität. Werden beispielsweise sämtliche GPS-Sensordaten herausgefiltert, so können von da an keine standortbezogenen Dienste mehr angeboten werden. Selbst wenn ein Aktivierungskontext genutzt wird, sind diese Dienste immer dann dysfunktional, wenn der Filter aktiv ist.

Musterbasierte Verfahren: Aus diesem Grund führen Stach et al. [St18] ein musterbasiertes Verfahren zum Schutz privater Daten ein. Das Ziel dabei ist, privaten Informationen vor einer IoT-Anwendung zu schützen, ohne die Servicequalität unnötig einzuschränken. Zu diesem Zweck werden auf die Konzepte des *Complex Event Processings* (*CEP*) [EB09] zurückgegriffen. Im CEP werden nicht einzelne Sensorwerte betrachtet, sondern höherwertige Ereignisse, die durch eine Folge von Werten innerhalb eines Zeitfensters repräsentiert werden. Ein Beispiel für ein solches Ereignis ist „Nutzer kommt in ungeheizte Wohnung“, das sich aus „Temperatur ≤ 15 °C“ und „Nutzer nähert sich Smart Home“ zusammensetzt. Auf diese Weise definiert der Nutzer *private Muster*, die er nicht preisgeben möchte, sowie *öffentliche Muster*, die für die Servicequalität relevant sind. Durch eine zeitliche Umsortierung der Ereignisse innerhalb des Datenstroms werden die privaten Muster aufgebrochen und somit verborgen. Eine Qualitätsfunktion bewertet alle möglichen Permutationen. Hierbei wird die Anzahl der erfolgreich verborgenen privaten Muster der der *False Positives* (öffentliche Muster, die durch die Umsortierung entstanden sind) und der *False Negatives* (öffentliches Muster, die durch die Umsortierung verborgen werden) gegenübergestellt. Die Permutation mit der besten Bewertung wird anschließend auf den Datenstrom angewandt.

Es ist offensichtlich, dass aufgrund des musterbasierten Ansatzes dieses Verfahren erheblich weniger restriktiv ist. Dadurch ist es möglich, einer IoT-Anwendung Sensordaten weitestgehend unverändert zur Verfügung zu stellen und lediglich bei einigen wenigen den Zeitstempel zu manipulieren, wodurch die Servicequalität erhalten bleibt. Allerdings stellt dieser Ansatz kein Mittel gegen die unnötige Preisgabe von detaillierten Sensordaten dar. Wenn beispielsweise ein öffentliches Muster „Temperatur ≤ 15 °C“ lautet, müsste der Anwendung die exakte Temperatur nicht bekannt sein – eine binäre Aussage („erfüllt“ oder „nicht erfüllt“) wäre ausreichend. Dies wird jedoch bei den Mustern nicht berücksichtigt.

Statistische Verfahren: Auch *Differential Privacy* findet im IoT-Kontext Anwendung. Birman et al. [Bi15] stellen einen solchen Ansatz für *Smart Grids* vor. Hierbei verbleiben die Daten vollständig auf den *Smart Metern* der Nutzer, während das Versorgungsunternehmen nur aggregierte Daten erhält. Differential-Privacy-Techniken stellen sicher, dass diese Daten keine Rückschlüsse auf einen individuellen Smart Meter (d. h. auf einen Nutzer) zulassen, die Datenqualität für eine statistische Auswertung dennoch maximal ist. Eine derartige Anonymisierung lässt sich jedoch nur in solchen Szenarien anwenden, in denen ausschließlich statistische Informationen über eine Menge an Nutzern, nicht aber über einen individuellen Nutzer benötigt werden. Für Anwendungsfälle, wie das in Abschnitt 2 beschriebene Smart-Home-Szenario, eignet sich dieses Vorgehen daher nicht. Hier muss es möglich sein, einen individuellen Nutzer exakt zu bestimmen, um zu entscheiden, welcher seiner Aktuatoren wie auf die Daten eines bestimmten Sensors reagieren soll.



(a) Unveränderte Rohdaten.

(b) Nullwerte als Mock-Daten.

(c) Lineare Interpolation.

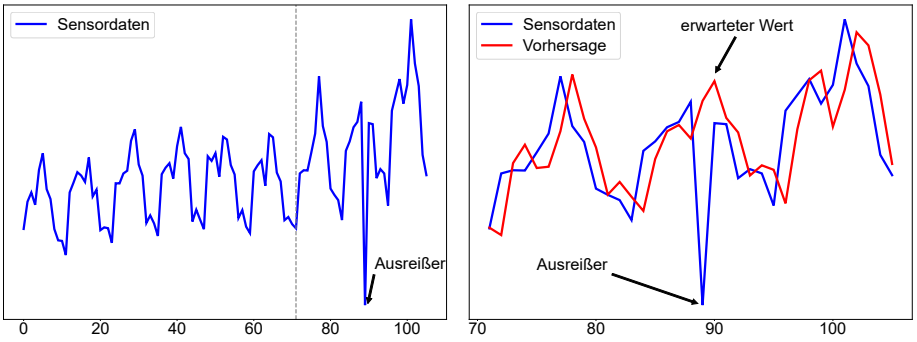
(d) Spline-Interpolation.

Abb. 2: Anwendungsbeispiel für die *Dateninterpolation* und die *Spline-Interpolation*.

4 Datenschutzkonzepte für Zeitreihendaten

Wie die Diskussion der verwandten Arbeiten zeigt, sind diese für die IoT-Domäne entweder zu restriktiv (attributbasierte Verfahren), geben unnötig viele Details weiter (Zugriffskontrollverfahren und musterbasierte Verfahren) oder lassen sich nicht für personenspezifische Anwendungsfälle nutzen (statistische Verfahren). Vielmehr sollte man für diesen Anwendungsbereich die charakteristischen Eigenschaften der Daten sowie der Verarbeitungsarten berücksichtigen. So besitzen bei Zeitreihendaten beispielsweise Ausreißer oft einen höheren Informationsgehalt und sollten daher besonders geschützt werden. Außerdem kommen bei der Verarbeitung dieser Daten häufig Kompressionstechniken zum Einsatz, um der schieren Datenmenge Herr zu werden. Wir stellen im Folgenden sechs Datenschutzkonzepte vor, die auf den Umgang mit Zeitreihendaten ausgelegt sind, d. h. deren Charakteristika ausnutzen.

Dateninterpolation: Datenpunkte, die stark von der Norm abweichen, besitzen oft einen wesentlich höheren Informationsgehalt, da sie auf ein ungewöhnliches Nutzerverhalten schließen lassen. Ein Beispiel hierfür ist in Abb. 2a gegeben. In der Kurve sei die Raumlautstärke über die Zeit aufgetragen. Wie man sieht, folgt diese einem zyklischen Muster (z. B. höhere Lautstärke am Tag als in der Nacht). In der Zeit zwischen $t_1 = 1.250$ und $t_2 = 1.570$ steigt der Lautstärkepegel jedoch unerwartet stark an. Aus diesem Werteverlauf könnte ein Angreifer beispielsweise ableiten, dass zu dieser Zeit mehr Personen als üblich in dem Raum anwesend waren. Möchte der Nutzer dies verbergen, könnte er die kompromittierenden Daten



(a) Unveränderte Rohdaten. (b) Vergleich von Ist-Werten und Vorhersage.

Abb. 3: Anwendungsbeispiel für die *Ausreißererkennung mittels Vorhersagen*.

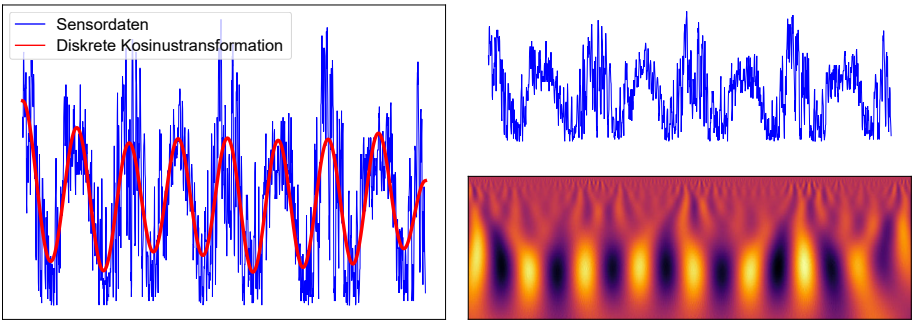
löschen. Dies würde allerdings zu Lücken im Werteverlauf führen und der Angreifer könnte erkennen, dass ihm Informationen vorenthalten wurden, und über den Grund spekulieren.

Ansätze wie *SmarPer* [Ol17] gestatten daher Sensordaten durch *Mock-Daten* (d. h. verfälschte Daten) zu ersetzen, beispielsweise wenn ein bestimmter Grenzwert überschritten wurde. Oft wird hierfür jedoch auf Nullwerte zurückgegriffen (siehe Abb. 2b). Für Zeitreihendaten ist dies allerdings aufgrund des atypischen, rapiden Wertabfalls ebenfalls sehr auffällig.

Wir schlagen daher vor, *lineare Interpolation* hierfür zu nutzen. Zu diesem Zweck reicht es aus, sich die beiden Datenpunkte vor und nach dem zu verschleiernden Bereich zu betrachten und damit die Gleichung $f(t) = y_1 * \frac{t_2-t}{t_2-t_1} + y_2 * \frac{t-t_1}{t_2-t_1}$ aufzulösen (siehe Abb. 2c). Dieses sehr einfache Verfahren führt allerdings zu kantigen Übergängen. Um noch glaubhaftere Mock-Daten zu erzeugen, kann auch auf *Spline-Interpolation* zurückgegriffen werden. Diese berücksichtigen auch die Steigungen am Anfang und Ende des zu füllenden Bereichs, was zu glatteren Übergängen führt (siehe Abb. 2d).

Ausreißererkennung mittels Vorhersagen: Während im obigen Beispiel davon ausgegangen wurde, dass Datenpunkte mit einem hohen Informationsgehalt einfach zu identifizieren sind (z. B. mittels eines Grenzwerts), kann sich dies im Allgemeinen als schwierig erweisen. Zu diesem Zweck kann eine Ausreißererkennung mittels Vorhersagen durchgeführt werden. Betrachtet man den Werteverlauf aus Abb. 3a, so kommt es zum Zeitpunkt $t_1 = 89$ zu einem Ausreißer, also einem Datenpunkt mit potentiell hohem Informationsgehalt. Dieser ist aber nicht offensichtlich, da es z. B. auch zum Zeitpunkt $t_2 = 101$ zu einem Extremum kommt.

In diesem Fall schlagen wir vor, mittels *maschinellen Lernens* eine Prognose für den weiteren Verlauf der Sensorwerte zu erstellen. Hierfür werden historische Daten des Nutzers als Trainingsdaten verwendet, um ein Modell zu lernen. Das heißt, die Werteverläufe der Vergangenheit werden analysiert und es wird eine *Hypothesenfunktion* $h(t)$ aufgestellt, die beschreibt, welcher Sensorwert zum Zeitpunkt t_{n+1} bei einem gegebenen $h(t_n)$ zu



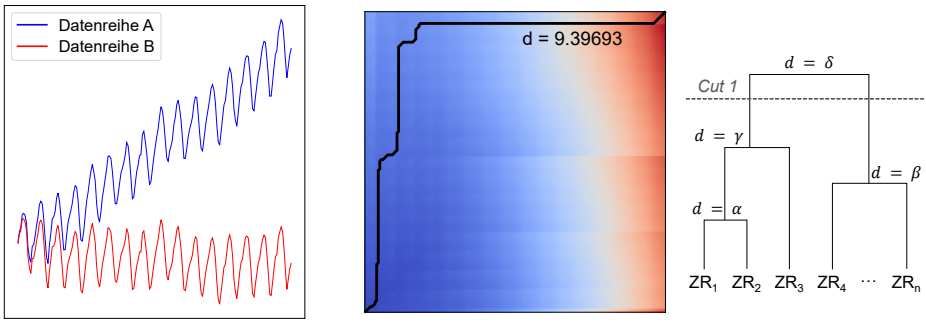
(a) Fourier-Transformation. (b) Wavelet-Transformation.

Abb. 4: Komprimierung von Zeitreihendaten durch *Fourier-* und *Wavelet-Transformationen*.

erwarten ist. Wir verwenden hierfür das *ARIMA-Modell*, das speziell für Zeitreihendaten ausgelegt ist, da es anstelle des vorherigen Sensorwerts das *gleitende Mittel* (Mittelwert der vorangegangenen n Werte) für die Prognose heranzieht. Die Aufteilung in Trainings- und Anwendungsdaten erfolgt zum Zeitpunkt $t_{split} = 71$, d. h. ab diesem Zeitpunkt erfolgt die Vorhersage. Abb. 3b stellt die Ist-Werte und die vorhergesagten Werte einander gegenüber. Weicht der Ist-Wert mehr als ein gegebenes Delta Δ_p von der Prognose ab, so handelt es sich nach unserer Definition um einen Ausreißer, d. h. um ein von der Norm abweichendes Verhalten, das nicht preisgegeben werden darf. Dieses Verfahren liefert glaubwürdige Mock-Daten von Haus aus mit, da im Fall von Ausreißern die prognostizierten Werte hierfür verwendet werden können. Der leichte Versatz der beiden Kurven ergibt sich durch das gleitende Mittel und kann durch Parameteranpassungen weiter verringert werden.

Signalglättung: Je nach Anwendung kann es aber auch erforderlich sein, genau diese Abweichungen von der Norm zu identifizieren. Im Rahmen von *Active Assisted Living* will man beispielsweise automatisch feststellen, ob ein Nutzer gestürzt oder anderweitig in eine Notsituation geraten ist. Hierzu muss gezielt nach ungewöhnlichen Mustern in den Daten gesucht werden, während die exakten Sensorwerte hingegen oft vernachlässigt werden können. Beispielsweise müssen die Daten keine exakten Bewegungsabläufe wiedergeben, wenn es für die Pflegekraft ausreicht, ungefähr zu wissen, wo der Sturz stattgefunden hat.

Zu diesem Zweck schlagen wir eine *Fourier-Transformation* vor. Hierbei werden zunächst die zeitdiskreten, äquidistanten Sensordaten von dem Zeitbereich auf ihren Frequenzbereich abgebildet, indem die Eingangssignalfolge als eine Summe von trigonometrischen Funktionen ausgedrückt wird. Wendet man diese Transformation auf aufeinanderfolgende Zeitfenster an, so erzielt man dabei einen *Bandfiltereffekt*, d. h. bei einer geeigneten Wahl des Zeitfensters lassen sich bestimmte Frequenzen abschwächen. Durch eine anschließende Anwendung der *inversen Fourier-Transformation*, erhält man eine geglättete Version der ursprünglichen Zeitreihe. In Abb. 4a ist dieser Effekt zu erkennen. Hierbei kommt die *diskrete Kosinustransformation* zum Einsatz. Die N Sensordaten $x_0, \ldots, x_{N-1}$ werden dabei auf die Frequenzen



(a) Anwendung des Dynamic-Time-Warping-Algorithmus auf Zeitreihendaten. (b) Clustering.

Abb. 5: Anwendungsbeispiel für eine *Clusteranalyse* von Zeitreihendaten.

$X_0, \ldots, X_{N-1}$ mittels folgender Funktion abgebildet: $X_k = \sum_{n=0}^{N-1} x_n * \cos\left[\frac{\pi}{N} * \left(n + \frac{1}{2}\right) * k\right]$. Die Werteverläufe sind weiterhin sichtbar, die Glättung eliminiert jedoch sämtliche Details.

Informationsexploration: Manche Anwendungen kommen mit noch weniger Daten aus. Eine Einbruchserkennung muss beispielsweise nur erkennen, dass ein ungewöhnliches Verhalten vorliegt. Daten über den Normalzustand werden hingegen überhaupt nicht benötigt.

In diesem Fall schlagen wir eine *Wavelet-Transformation* vor. Dabei handelt es sich um eine alternative Zeit-Frequenz-Transformation. Im Gegensatz zur Fourier-Transformation, die eine gleichbleibende Zeitfenstergröße für die Transformation nutzt, skaliert die Wavelet-Transformation das Zeitfenster. Auf diese Weise wird abhängig vom Eingangssignal eine bessere zeitliche Auflösung oder Frequenzauflösung erzielt, was insbesondere bei abrupten Frequenzwechseln zu besseren Ergebnissen führt. In Abb. 4b wird das *Gaußsche Wellenpaket* als *Wavelet* auf das abgebildete Signal angewandt. In der grafischen Ergebnisrepräsentation darunter ist klar ersichtlich, dass dadurch Anomalien in den Daten betont werden (Maxima – hell und Minima – dunkel). Diese können im Anschluss der Anwendung gemeldet werden.

Clusteranalyse: Wurden die Anomalien entdeckt, so können diese gruppiert werden. Hierzu empfehlen wir ein *Cluster-basiertes* Verfahren. Zunächst werden die Zeitreihendaten in Abschnitte unterteilt (ein Abschnitt je Anomalie). Diese Abschnitte werden miteinander verglichen, indem paarweise ein Distanzmaß ermittelt wird. In Abb. 5a wird hierfür der *Dynamic-Time-Warping-Algorithmus* [RK04] verwendet. Bei diesem Algorithmus werden zwei unterschiedliche Signale (Datenreihe A und B) aufeinander abgebildet. Mittels unterschiedlicher Transformationen werden die einzelnen Datenpunkte des einen in die des anderen Signals überführt. Jede Transformation ist mit spezifischen Kosten verbunden. Diese Kosten werden in eine Matrix eingetragen. Der Algorithmus findet den *kürzesten* (i. S. v. kostengünstigsten) Weg von dem einen ins andere Signal (siehe Abb. 5a, rechte Seite). Die Kosten für diesen Weg stehen für die Distanz der beiden Signale. Für die Gruppierung stellt jeder Abschnitt stellt zunächst ein eigenes Cluster dar (siehe Abb. 5b). Schrittweise

A	B	C	D
α	β	γ	δ
ε	ζ	η	θ
ι	κ	λ	μ

(a) Ausgangsdaten.

A	C	D
α	γ	δ
ε	η	θ
ι	λ	μ

(b) Projektion.

A	B	C	D
α	β	γ	δ
ι	κ	λ	μ

(c) Selektion.

Abb. 6: Auswirkungen einer *Projektion* und einer *Selektion* auf eine Datenbankanfrage.

werden anschließend die Cluster mit der kleinsten Distanz zusammengefasst, bis alle in einem einzigen Cluster liegen. Abhängig davon, wie viele Gruppen (d. h. Anomalie-Typen) man unterscheiden möchte, trennt man die Hierarchie auf (*agglomerative hierarchische Clusteranalyse*). In Abb. 5b werden zwei Gruppen von Anomalien gebildet (*Cut 1*). Diese können mit einem sprechenden Schlagwort versehen werden, das das zugrundeliegende Ereignis beschreibt (z. B. „Sturz“). Anschließend können die von Stach et al. [St18] vorgestellten Techniken angewandt werden, um deren zeitliche Abfolge zu verschleiern.

Query Rewriting: Selbst wenn die Daten bereits der Batch-Verarbeitung vorliegen und keine privaten Informationen entfernt wurden, können die Daten noch geschützt werden. Hierfür kann *Query Rewriting* genutzt werden, d. h. eine Anfrage wird automatisch vor der Ausführung umgeschrieben. List. 1 verdeutlicht dieses Vorgehen. Stellt der Nutzer beispielsweise die Anfrage a an eine Datenbank, so erhält er alle in der Tabelle „meine_sensoren“ gespeicherten Daten. Eine *Projektion* blendet bestimmte Spalten (d. h. Sensorattribute) aus. Anfrage b sorgt dafür, dass der Nutzer nur noch Temperaturdaten erhält. Eine *Selektion* blendet hingegen Spalten (d. h. Zeitbereiche) aus. Anfrage c gibt nur die Daten der letzten Woche zurück. Die Auswirkungen dieser Manipulationen sind in Abb. 6 veranschaulicht. Schließlich kann eine *Aggregation* die Daten zusätzlich verdichten. So gibt Anfrage d nur die durchschnittliche Temperatur je Raum zurück.

Diskussion der Datenschutzkonzepte: Die sechs vorgestellten Datenschutzkonzepte für Zeitreihendaten operieren auf unterschiedlichen Abstraktionsleveln. Je nach Anwendungsfall kann eine andere Strategie zum Einsatz kommen. Dadurch kann gewährleistet werden, dass private Informationen geschützt werden und dennoch die Servicequalität

```
1 SELECT *
2 FROM "meine_sensoren"
```

(a) Anfrage über alle Daten.

```
1 SELECT "temperatur"
2 FROM "meine_sensoren"
```

(b) Anwendung einer Projektion.

```
1 SELECT *
2 FROM "meine_sensoren"
3 WHERE time > now() - 7d
```

(c) Anwendung einer Selektion.

```
1 SELECT MEAN("temperatur")
2 FROM "meine_sensoren"
3 GROUP BY "raum"
```

(d) Anwendung einer Aggregation.

List. 1: Anwendungsbeispiele für *Query Rewriting*.

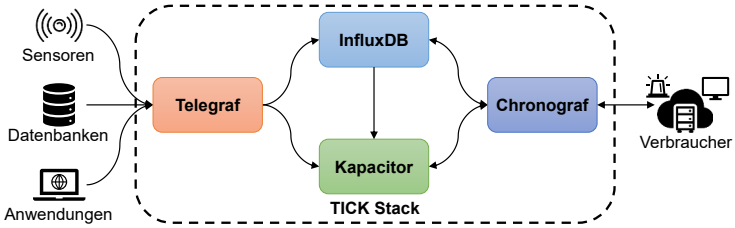


Abb. 7: Der *TICK-Stack* als Implementierung der *Lambda-Architektur*.

aufrechterhalten bleibt, d. h. die IoT-Anwendung wird weiterhin mit ausreichend Daten versorgt, um ihre Analysen durchzuführen. Mittels **Dateninterpolation** können konkrete Werte(bereiche) mit einem hohen Informationsgehalt verborgen werden. Abhängig davon, wie wichtig es ist, dass einem Angreifer diese Manipulation nicht auffällt, kann hierfür beispielsweise eine einfache lineare Interpolation, Splines oder maschinelles Lernen genutzt werden. Die **Vorhersagen**, die letzteres ermöglicht, helfen auch bei der Identifikation von potentiell schützenswerten Datenpunkten. Sollen nicht nur einzelne Datenpunkte verborgen werden, sondern eine Unschärfe auf das gesamte Sensorsignal angewandt werden, eignen sich Fourier-Transformationen zur **Signalglättung**. Die Wavelet-Transformationen führt diese Filterung noch weiter, indem sie Anomalien in den Daten pointiert. Dies dient der **Informationsexploration**. Auf diesen Anomalien kann eine **Clusteranalyse** durchgeführt werden, um die Daten noch weiter zu verdichten. Auch Kombinationen dieser Strategien sind möglich. So können beispielsweise alle Anomalien identifiziert werden und ausgewählte mittels Dateninterpolation verborgen werden. Während diese Strategien proaktiv sind, schützt **Query Rewriting** die Daten auch, wenn sie bereits an die Batch-Verarbeitung weitergeleitet wurden. Dieses Verfahren ist sehr mächtig und kann beliebige Restriktionen auf Anfragen anwenden. Allerdings steigt die Komplexität dieses Verfahrens je nach Restriktion stark an.

Im Nachfolgenden stellen wir Implementierungsansätze für diese Schutzkonzepte vor und diskutieren anhand des TICK-Stacks, wie dadurch *Privacy by Design* ermöglicht wird.

5 Implementierungsansätze

Der TICK-Stack ist eine Implementierung der Lambda-Architektur (siehe Abb. 7). Er besteht aus vier Komponenten: Der *Telegraf* (Verteilerkomponente) sammelt Daten von beliebigen Datenquellen und stellt diese der *InfluxDB* (Batch-Verarbeitung) und dem *Kapacitor* (Datenstromverarbeitung) zur Verfügung. Die InfluxDB ist eine dedizierte Datenbank für Zeitreihendaten während der Kapacitor eine Verarbeitungs-Engine für Zeitreihendaten ist. Neben den Daten von Telegraf hat er auch Zugriff auf die InfluxDB. Der *Chronograf* kann mit diesen beiden Komponenten interagieren und die dort berechneten Analyseergebnisse für Verbraucher aufbereiten. Er stellt gleichzeitig die Schnittstelle zum Nutzer dar, der darüber Anfragen an die InfluxDB und den Kapacitor stellen kann. Somit besitzt der TICK-Stack alle Funktionalitäten, die eine IoT-Anwendung benötigt (siehe Abschnitt 2).

Wir haben die in Abschnitt 4 vorgestellten Konzepte als *Python-Skripte* umgesetzt, die selbst auf einem *Raspberry Pi* effizient ausgeführt werden können. Somit könnten sie beispielsweise in der *Edge*, also an einem vom Nutzer kontrollierten Punkt, die Daten vorfiltern, bevor diese zur Verarbeitung weitergereicht werden. Da die einzelnen Komponenten des TICK-Stacks verteilt ausgeführt werden können, ist es daher möglich den Telegraf ebenfalls in der Edge zu betreiben und darin unsere Skripte zu integrieren. Somit ist sichergestellt, dass die InfluxDB und der Kapacitor keine privaten Daten mehr erhalten. Da diese beiden Komponenten aber selbst ebenfalls Python-Schnittstellen besitzen, können unsere Skripte auch anbieterseitig angewandt werden, um so Privacy by Design zu realisieren. Ein Beispiel hierfür wäre eine rollenbasierte Zugriffsschicht, die dem Chronograf vorgeschaltet ist. Je nach IoT-Anwendung wählt diese Zugriffsschicht die entsprechenden Datenschutzkonzepte aus (entsprechend der Rolle der Anwendung), um dadurch eine DSGVO-konforme Verarbeitung zu gewährleisten.

6 Zusammenfassung

Die massiven Fortschritte, die IoT-fähige Geräte in den letzten Jahren bezüglich Rechenleistung, Übertragungsgeschwindigkeit sowie Sensorik erfahren haben, haben die technischen Voraussetzungen für eine Vielzahl an IoT-Anwendungen geschaffen. Sie durchdringen sämtliche Bereiche des täglichen Lebens, beispielsweise Smart Homes, den Gesundheitssektor oder die Industrie 4.0. Um die Vorteile dieser Anwendungen genießen zu können, müssen Nutzer allerdings viele und zum Teil hochgradig private Daten preisgeben. Heutige Datenschutzlösungen sind allerdings nicht auf die speziellen Eigenschaften der hierbei zum Einsatz kommenden Zeitreihendaten angepasst, wodurch sie unnötig restriktiv sind.

In diesem Artikel stellen wir daher sechs Datenschutzkonzepte speziell für Zeitreihendaten vor. Diese Konzepte berücksichtigen neben der Struktur der Daten auch deren spätere Verarbeitungsarten. So ist die *Dateninterpolation* und die *Ausreißererkennung* vergleichbar mit der Rauschunterdrückung, die auf Sensordaten häufig angewandt wird, während die *Signalglättung* und *Informationsexploration* respektive *Clusteranalyse* oft zur Komprimierung von Sensordaten genutzt werden. Mithilfe von *Query Rewriting* kann der Zugriff auf die Daten weiter eingeschränkt werden. Auf diese Weise lassen sich private Daten zurückhalten und damit gezielt private Informationen vor IoT-Anwendungen verbergen. Gleichzeitig wird die Servicequalität beinahe beibehalten, da viele der Operationen ohnehin zu einem späteren Zeitpunkt angewandt worden wären. Diese Konzepte haben wir als leichtgewichtige Python-Skripte umgesetzt, die sich beispielsweise in den TICK-Stack integrieren lassen. Damit kann die Forderung der DSGVO nach einer handhabbaren Privacy-by-Design-Datenschutzlösung erfüllt werden, ohne die Servicequalität von IoT-Anwendungen unnötig einzuschränken. In zukünftigen Arbeiten gilt es nun zu prüfen, wie sich die Konfiguration der Skripte (d. h. die Spezifikation der Privacy-Richtlinien) nutzerspezifisch automatisieren lässt (z. B. mithilfe von maschinellem Lernen), um so die Belastung für den Nutzer zu reduzieren.

Danksagung: Die in diesem Beitrag vorgestellte Forschungsarbeit entstand aus dem PATRON-Forschungsauftrag, der von der Baden-Württemberg Stiftung finanziert wurde.

Literatur

[Ad11] Adaikkalavan, R. et al.: Multilevel Secure Data Stream Processing. In: DBSec '11. 2011.

[Bh19] Bhattacharjya, A. et al.: Security Challenges and Concerns of Internet of Things (IoT). In: Cyber-Physical Systems. Springer, Cham, Kap. 7, S. 153–185, 2019.

[Bi15] Birman, K. et al.: Building a Secure and Privacy-Preserving Smart Grid. SIGOPS Oper. Syst. Rev. 49/1, S. 131–136, 2015.

[Bo18] Bourgeois, J. et al.: Trusted and GDPR-compliant Research with the Internet of Things. In: IOT '18. 2018.

[EB09] Eckert, M.; Bry, F.: Complex Event Processing (CEP). Informatik-Spektrum 32/2, S. 163–167, 2009.

[Gr19] Gritti, C. et al.: Privacy-Preserving Delegable Authentication in the Internet of Things. In: SAC '19. 2019.

[Ma17] Marjani, M. et al.: Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges. IEEE Access 5/1, S. 5247–5261, 2017.

[Ma19] Marikyan, D. et al.: A systematic review of the smart home literature: A user perspective. Technol. Forecasting Social Change 138/1, S. 139–154, 2019.

[MW15] Marz, N.; Warren, J.: Big Data: Principles and best practices of scalable real-time data systems. Manning Publications Co., Shelter Island, NY, 2015.

[Ol17] Olejnik, K. et al.: SmarPer: Context-Aware and Automatic Runtime-Permissions for Mobile Devices. In: S&P '17. 2017.

[Ou16] Ouaddah, A. et al.: Access control in IoT: Survey & state of the art. In: ICMCS '16. 2016.

[PO12] Pieterse, H.; Olivier, M.: Data Hiding Techniques for Database Environments. In: DigitalForensics '12. 2012.

[RK04] Ratanamahatana, C. A.; Keogh, E.: Everything you know about Dynamic Time Warping is Wrong. In: TDM '04. 2004.

[St18] Stach, C. et al.: How a Pattern-based Privacy System Contributes to Improve Context Recognition. In: PERCOM WKSHPS '18. 2018.

[Ve17] Velosa, A. et al.: Hype Cycle for the Internet of Things, 2018, Report G00340237, Gartner, 17. Juli 2017.

[Vo17] Voss, W. G.: First the GDPR, Now the Proposed ePrivacy Regulation. Journal of Internet Law 21/1, S. 3–11, 2017.

[Wa18] Wachter, S.: Normative challenges of identification in the Internet of Things. Computer Law & Security Review 34/3, S. 436–449, 2018.

[Zh18] Zheng, S. et al.: User Perceptions of Smart Home IoT Privacy. Proc. ACM Hum.-Comput. Interact. 2/CSCW, 200:1–200:20, 2018.

Datenqualität bei algorithmischen Entscheidungen

Jeremy Stevens[1]

Abstract: Die Kontrolle der Ergebnisse algorithmischer Entscheidungen auf Basis von Machine Learning-Verfahren stellt das Recht vor eine Vielzahl von Herausforderungen. Entsprechende Algorithmen analysieren umfangreiche Datenbestände und nutzen die in den Daten enthaltenen Muster und Gesetzmäßigkeiten zur Prognose. Die Qualität der verwendeten Daten hat daher einen erheblichen Einfluss auf das Ergebnis algorithmischer Entscheidungen. Der Beitrag soll aufzeigen, dass bestehende datenschutzrechtliche Konzepte zur Überprüfung der Qualität der verwendeten Daten keine hinreichende Kontrolle gewährleisten. Weder das System individueller Betroffenenrechte noch die Befugnisse der Datenschutzbehörden sind hierfür ausreichend. Dementsprechend sollen zuletzt Regulierungsmöglichkeiten aufgezeigt werden.

Keywords: Datenqualität; Algorithmische Entscheidungen; DSGVO

1 Einleitung

Künstliche Intelligenz und deren Anwendung im Rahmen automatisierter Entscheidungen stellen das Recht vor eine Vielzahl von Herausforderungen, insbesondere vor die Frage, inwiefern eine Kontrolle algorithmischer Entscheidungsergebnisse gewährleistet werden kann. Die fortschreitende technische Entwicklung macht dabei zunehmend komplexe Anwendungen und eine Automatisierung weiter Lebensbereiche möglich. Das Versprechen einer objektiven und präzisen Bearbeitung von Massenverfahren macht algorithmische Entscheidungsverfahren für Wirtschaft und Verwaltung interessant.[2] Berichte über bestehende Anwendungen bieten jedoch einige problematische Beispiele, in denen der Einsatz von Algorithmen zu falschen oder diskriminierenden Ergebnissen geführt hat. So besteht berechtigter Anlass zur Sorge, dass der im US-Justizsystem eingesetzte COMPAS-Algorithmus zur Einschätzung der Rückfallwahrscheinlichkeit verurteilter Straftäter erheblich zulasten afroamerikanischer Personen urteilt.[3] Ähnliche Bedenken bestehen im Rahmen des polizeilichen Einsatzes von Algorithmen zur Prognose

[1] Goethe-Universität Frankfurt, Lehrstuhl für Öffentliches Recht, Informationsrecht, Umweltrecht, Verwaltungswissenschaft, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt am Main, stevens@jur.uni-frankfurt.de
[2] Entsprechende Überlegungen finden sich beispielsweise in der KI-Strategie der Bundesregierung.
[3] Siehe hierzu die Recherche von ProPublica, verfügbar unter https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing; https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm

doi:10.18420/inf2019_55

von Straftaten.[4] Mehrere Projekte zum Einsatz von IBM Watson zur Behandlung von Krebspatienten wurden wieder eingestellt, nachdem der Algorithmus häufig fehlerhafte Diagnosen und Behandlungsempfehlungen erstellt hatte.[5] Ziel einer rechtlichen Kontrolle algorithmischer Entscheidungen muss sein, solche Fehleinschätzungen zu vermeiden.

Moderne Algorithmen basieren auf Machine Learning-Methoden. Diese erstellen auf Grundlage einer Analyse umfangreicher Datenbestände eigenständige Entscheidungsmodelle. Eine effektive Regulierung algorithmischer Entscheidungsprozesse setzt ein Verständnis dieser Methoden voraus. Der Beitrag beginnt daher mit einer kurzen Darstellung der Grundlagen des Machine Learning mit einem besonderen Fokus auf die Bedeutung der verwendeten Daten. Im Anschluss werden bestehende Datenqualitätsstandards im Datenschutzrecht analysiert und Probleme mit deren Anwendung auf algorithmischen Entscheidungssystemen aufgezeigt. Zuletzt sollen alternative Regulierungsmöglichkeiten skizziert werden.

2 Grundlagen des Machine Learning und Bedeutung der Datenqualität

Der Begriff des Machine Learning umfasst verschiedene statistische Methoden, durch die Algorithmen ohne umfassende menschliche Programmierung Lösungsmodelle für spezifische Aufgaben entwickeln können. Machine Learning Algorithmen nutzen umfangreiche Datenbestände als Input, um mittels statistischer Analyse Muster und Gesetzmäßigkeiten in den Daten zu erkennen und auf dieser Grundlage einen entsprechenden Output zu prognostizieren. Die Funktionsweise der einzelnen Methoden weicht zum Teil erheblich voneinander ab. Grundlegende Strukturen des Entwicklungsprozesses lassen sich aber für die rechtliche Diskussion verallgemeinern.

Ausgangspunkt der Entwicklung eines Machine Learning Algorithmus sind zumeist umfangreiche Datenbestände. Hierbei handelt es sich in der Regel um Big Data, daher große Mengen von unstrukturierten und semi-strukturierten Daten.[6] Diese Daten können unmittelbar bei den Betroffenen erhoben werden, auf den Angaben und Einschätzungen von Mitarbeitern und Einsatzkräften, wie medizinischem Personal und Polizeibeamten, beruhen oder aus bestehenden Datenbanken bezogen werden.

Algorithmen prüfen diese Datenbestände im Rahmen einer statistischen Analyse auf Korrelationen zwischen einzelnen Inputvariablen und dem zu prognostizierenden Output. Die hierbei gefundenen Muster dienen als Grundlage des Prognosemodells. Aufgrund der häufig hohen Anzahl an Variablen und der Funktionsweise vieler Methoden ist

---

[4] Siehe hierzu Ki17.

[5] https://www.faz.net/aktuell/wirtschaft/kuenstliche-intelligenz/computer-watson-scheitert-zu-oft-bei-datenanalyse-15619989.html

[6] Daneben beschreibt der Begriff auch Technologien, welche zur Verarbeitung der umfangreichen Datenmengen eingesetzt werden.

die Herleitung der Prognose für Menschen häufig gar nicht oder nur mit erheblichem Aufwand nachvollziehbar. Aus diesem Grund wird im Zusammenhang mit Machine Learning häufig von einer Black Box gesprochen.[7] Die fehlende Interpretierbarkeit kann durch aktuelle Methoden zur Erklärbarkeit algorithmischer Ergebnisse auch nicht hinreichend ausgeglichen werden.[8]

Nach Abschluss der Entwicklung soll der Algorithmus im Realeinsatz zutreffende Prognosen erstellen können. Eine fehlerhafte Datengrundlage oder die Anwendung in einem abweichenden Kontext können zu fehlerhaften Entscheidungen führen, ohne dass dies für Betroffene erkennbar wäre. Ein Bericht der Investigativplattform ProPublica legt eine solche Situation im Fall des COMPAS-Algorithmus nahe.[9] Die Veröffentlichung hat eine umfassende Diskussion über den Einsatz und die Beurteilung von Prognosesoftware im Justizsystem ausgelöst.[10] Die fehlerhaften Prognosen von IBM Watson lassen sich wiederum zum Teil auch auf eine Diskrepanz zwischen den in der Entwicklung genutzten Datenbeständen und der Anwendung in den einzelnen Krankenhäusern zurückführen. Die Entwicklung erfolgte in Kooperation mit einem US-Krankenhaus mit überwiegend wohlhabenden Patienten. Muster in den Krankheitsbildern bilden diese spezifischen Umstände ab und sind als Grundlage einer Prognose lediglich für diese spezielle Situation geeignet. Eine Übertragung des Algorithmus auf andere Krankenhäuser führt zu fehlerhaften Ergebnissen.[11]

Die verwendeten Daten haben einen erheblichen Einfluss auf algorithmische Entscheidungen. Die Gewährleistung einer qualitativen Datengrundlage stellt einen wichtigen Bestandteil der Kontrolle algorithmischer Entscheidungen dar. Wie die Beispiele zeigen, handelt es sich bei der Beurteilung der Datenqualität jedoch um eine komplexe Aufgabe, bei der unter anderem der Anwendungskontext, mögliche Fehler oder Verzerrungen bei der Datenerhebung oder strukturelle Benachteiligung einzelner Bevölkerungsgruppen relevant sein können.

Fraglich ist vor diesem Hintergrund, wie das Recht mit der Problematik der Datenqualität umgehen soll.

3 Rechtliche Anforderungen an die Datenqualität

Im Folgenden sollen kurz bestehende rechtliche Anforderungen an die Datenqualität, insbesondere im Zusammenhang mit algorithmischen Entscheidungen, untersucht wer-

---

[7] So beispielsweise Sh18, Kr17, Wi18, Pa15.

[8] Entsprechende Methoden, wie LIME (Local Interpretable Model-Agnostic Explanations, siehe hierzu RSG16) bieten zwar Ansätze zur Erklärbarkeit, sind aber nicht in jeder Situation wirksam einsetzbar.

[9] Eine abschließende Beurteilung ist bei kommerziellen Algorithmen häufig schwierig, da die Anbieter deren Funktionsweise als Geschäftsgeheimnis behandeln.

[10] Zum Bericht von ProPublica siehe Fn. 3; eine Übersicht über die Gesamtdebatte bietet Wa.

[11] Siehe Fn. 5

den. Als Untersuchungsgegenstand bietet sich insoweit die Datenschutzgrundverordnung an.[12] Diese befasst sich zum einen mit der Regulierung algorithmischer Entscheidungen[13] und enthält zum anderen Vorgaben zur Datenqualität, die bisher noch nicht so viel Beachtung gefunden haben.[14] Daneben soll der Umgang mit Datenqualität im US-Recht beleuchtet und der rechtlichen Sichtweise ein Überblick über technische Konzeptionen der Datenqualität entgegengestellt werden.

3.1. Anforderungen der Datenschutzgrundverordnung

Vorrangiger Anknüpfungspunkt im Rahmen der Datenschutzgrundverordnung ist Art. 5 Abs. 1 lit. d DSGVO, der die sachliche Richtigkeit der Daten zu einem Grundsatz der Datenverarbeitung erhebt[15]. Daneben könnten auch die Anforderungen an den Einsatz algorithmischer Entscheidungssysteme nach Art. 13 ff. und 22 DSGVO möglicherweise Auswirkungen auf Datenqualitätsstandards haben.

Art. 5 Abs. 1 lit. d DSGVO schreibt vor, dass personenbezogenen Daten „sachlich richtig und erforderlichenfalls auf dem neuesten Stand sein“ müssen und darüber hinaus „alle angemessenen Maßnahmen zu treffen [sind], damit personenbezogene Daten, die im Hinblick auf die Zwecke ihrer Verarbeitung unrichtig sind, unverzüglich gelöscht oder berichtigt werden („Richtigkeit”)“.

Die Auslegung des Begriffs der „Richtigkeit“ im Sinne von Art. 5 Abs. 1 lit. d DSGVO ist noch nicht eindeutig geklärt. Basierend auf dem deutschen Wortlautverständnis wird die Anwendung zum Teil auf Fakten im Sinne einer klaren Trennung zwischen „richtig“ und „unrichtig“ beschränkt.[16] Zwar kann hiernach auch eine unvollständige Darstellung der Wirklichkeit zur Unrichtigkeit der Daten führen. Insgesamt beschränkt sich dieser Ansatz jedoch auf eindeutig nachweisbare Daten. Ein Kontextbezug besteht insoweit lediglich hinsichtlich der Aktualität der Daten. Persönliche Daten sind grundsätzlich auf dem neusten Stand zu halten, es sei denn, sie beziehen sich explizit auf einen früheren Zeitpunkt.[17]

Die Gegenansicht beruft sich auf die komplexere Bedeutung des „accuracy“-Begriffs der englischen Sprachfassung[18] und befürwortet eine Ausweitung des Begriffs der Richtig-

[12] Wie Ho16a in Bezug auf Big Data zutreffend hinweist, betrifft die Frage der Datenqualität nicht nur personenbezogene Daten, sondern ebenso anonyme oder sachbezogene Daten.
[13] Dies betrifft insbesondere Art. 22 DSGVO, welcher „automatisierte Entscheidungen im Einzelfall“ regeln soll.
[14] Siehe insoweit Ho16a, Ho16b.
[15] Dieser ist gemäß Art. 83 Abs. 5 lit. a DSGVO mit einer erhebliche Bußgeldandrohung verbunden, weswegen die Norm auch in der Praxis von hoher Bedeutung sein wird.
[16] So Ro19.
[17] So Ro19.
[18] Der französische Wortlaut „exactitude“ ermöglicht eine vergleichbare Auslegung des Begriffs.

keit, im Sinne von Zielgerichtetheit oder Genauigkeit, auf Werturteile und Prognosen.[19] Eine unrichtige Prognose wäre nach dieser Ansicht beispielsweise bei Zugrundelegung falscher Tatsachen gegeben.[20]

Die Diskussion um die sachliche Richtigkeit von Daten wurde bisher vorwiegend im Rahmen der Betroffenenrechte nach Art. 12 ff. DSGVO geführt. [21] Bei der Beurteilung des Anspruchs auf Berichtigung unrichtiger Daten nach Art. 16 DSGVO und vergleichbaren Sachverhalten[22] sind insbesondere durch den Betroffenen beweisbare Tatsachen relevant. Bei der Kontrolle nichtrechtlicher Wertungen halten sich Gerichte mangels entsprechender Fachkompetenz in der Regel zurück.[23] Angesichts einfacher Sachverhalte ist eine enge Konzeption der Datenqualität möglicherweise ausreichend. In Bezug auf Big Data und algorithmische Entscheidungen sind einem solchen Ansatz jedoch Grenzen gesetzt. Big Data ist per Definition zu umfangreich und zu komplex, um diese mit manuellen Methoden zu bewältigen. Eine Kontrolle der Datenqualität muss daher gegebenenfalls an anderen Kriterien anknüpfen können. Dies gilt insbesondere bei Daten, deren Richtigkeit nicht ohne Weiteres ersichtlich ist sowie bei kontextbezogenen Daten als Grundlage einer algorithmischen Entscheidung.

Ein Beispiel hierfür sind Daten zur Social Media-Nutzung oder dem Einkaufsverhalten bestimmter Personen. Die einzelnen Likes bzw. Einkaufsartikel tragen für den Einzelnen häufig keine große Bedeutung. Der Betroffene wird daher häufig nicht in der Lage sein, die Daten vollständig zu verifizieren. Das gleiche gilt für auf dieser Grundlage getroffenen Prognosen über zukünftiges Verhalten. Die Probleme bei der Nutzung der medizinischen Prognosen des IBM Watson zeigen deutlich, dass Daten in einer spezifischen Situation inhaltlich zutreffend erhoben worden sein und trotzdem in einer anderen Situation zu unbrauchbaren Ergebnissen führen können.

Ein differenzierterer rechtlicher Datenqualitätsstandard im Sinne des „accuracy"-Ansatzes wäre eher geeignet, technische Standards zu integrieren und einen nutzbaren Maßstab für die Kontrolle von Big Data Anwendungen zu liefern.[24] Mangels einer eindeutigen Begriffsbedeutung verbleibt bis zur konkreten Klärung eine gewisse Rechtsunsicherheit bei der Anwendung des Grundsatzes.[25]

Weitere Anforderungen zur Datenqualität könnten gegebenenfalls aus den spezifischen Anforderungen an den Einsatz automatisierter Verfahren nach Art. 13 Abs. 2 lit. f, Art. 14 Abs. 2 lit. g, Art. 15 Abs. 1 lit. h DSGVO und Art. 22 DSGVO folgen. Demnach sind

---

[19] Ho16a; Scb Eine weite Anwendung liegt auch den Überlegungen der Art. 29 Datenschutzgruppe zugrunde, auch wenn dies nicht explizit benannt wird, siehe Ar18.
[20] Scb.
[21] Vergleiche die Diskussion in diesem Rahmen mit teils anderen Positionen: Ho16a, Wo.
[22] Vergleichbare Entscheidungen über die Richtigkeit von Aussagen sind beispielsweise in Unterlassungsklageverfahren zu treffen. In diesem Kontext schränkt die Meinungsfreiheit die Kontrolle wertender Äußerungen ein.
[23] Ein Beispiel hierfür bietet die umfassende Rechtsprechung zur Beurteilung von Prüfungsleistungen.
[24] So auch Ho16a.
[25] Deswegen für eine nachsichtige Anwendung plädierend Ho16a.

Betroffenen aussagekräftige Informationen über die involvierte Logik des automatisierten Verfahrens zur Verfügung zu stellen. Dies wird durch Erwägungsgrund 71 ergänzt, der die Verwendung geeigneter mathematischer oder statistischer Verfahren sowie technische und organisatorische Maßnahmen zur Sicherstellung richtiger Daten und Fehlervermeidung vorsieht. Der Begriff der „involvierten Logik" sowie die in Erwägungsgrund 71 angesprochenen Vorkehrung könnten bei entsprechender Auslegung zur Bestimmung notwendiger Datenqualitätsstandards herangezogen werden. Soweit ersichtlich werden aber bisher keine entsprechenden Konzepte in diesem Rahmen diskutiert. Vielmehr beschränkt die Literatur die Information zur involvierten Logik auf die Darstellung der entscheidungsrelevanten Kriterien. [26]

3.2. Datenqualität im US-Recht

Die Datenqualität ist auch Gegenstand des US-amerikanischen Rechts. Interessant sind insoweit insbesondere der Privacy Act 1974, der Data Quality Act 2001 und die hierzu ergangenen Richtlinien sowie der Fall Loomis v. Wisconsin.

§ 552a (e) (5) des Privacy Act 1974 verpflichtet staatliche Behörden bei Entscheidungen gegenüber Einzelpersonen „accuracy, relevance, timeliness, and completeness" der zugrunde gelegten Daten sicherzustellen. Anstelle eines allgemeinen Begriffs werden direkt mehrere Qualitätsmerkmale formuliert. Die Merkmale der Genauigkeit, Relevanz, Aktualität und Vollständigkeit der Daten werden in der Datenschutzgrundverordnung zwar ebenfalls erfasst. Die Formulierung des Privacy Act ist jedoch klarer gefasst und erlaubt eine differenzierte Anwendung einzelner Qualitätsmerkmale. Der Begriff der „accuracy" im Kontext des Privacy Act entspricht einem weiten Verständnis der Richtigkeit gemäß Art. 5 Abs. 1 lit. d DSGVO im Sinne von Genauigkeit und Zielgerichtetheit.[27] Anders als die Datenschutzgrundverordnung ist der Privacy Act jedoch auf die Datenverarbeitung von staatlichen Behörden beschränkt. Für andere Rechtsbereiche existieren vergleichbare Spezialgesetze.[28]

Im Rahmen des Data Quality Act[29] wurde das Office of Management and Budget zum Erlass v on Richtlinien ermächtigt, um „quality, objectivity, utility, and integrity of information" zu gewährleisten. Die auf dieser Grundlage erlassenen Vorgaben zur Daten- bzw. Informationsqualität sind vorwiegend verfahrensbezogen, wie beispielsweise die Gewährleistung der Objektivität durch Peer-Review-Verfahren.[30] Auch wenn diese Standards vorwiegend mit Blick auf die Veröffentlichung von Studien geschaffen wurden, lassen sich hieraus Schlussfolgerungen für rechtliche Anforderungen an die Daten-

[26] So beispielsweise Sca, Di19 und die Überlegungen der Art. 29 Datenschutzgruppe, siehe Ar18.
[27] Siehe hierzu beispielsweise Doe v. U.S., 821 F.2d 694 (C.A.D.C.,1987)
[28] Weitere Beispiele und Verweise bei Ho16a.
[29] Der Data Quality Act bzw. Information Quality Act wurde als sogenannter "rider" als Section 515 of the Consolidated Appropriations Act, 2001 durch den Kongress beschlossen und hat daher keinen offiziellen Namen.
[30] Zu einer Analyse dieses Ansatzes im Rahmen des Data Quality Acts, siehe Ga.

qualität algorithmischer Entscheidungen ziehen. Der Data Quality Act zeigt insbesondere auf, dass eine Regulierung der Daten- oder Informationsqualität sich auch auf deren Inhalt auswirkt.[31] Denn verschiedene Lobbyorganisationen haben versucht mit dem Vorwurf einer fehlenden Informationsqualität kritische Studien zu unterdrücken, beispielsweise zur gesundheitsschädlichen Wirkung übermäßigen Salzkonsums.[32]

Der Supreme Court of Wisconsin hatte im Verfahren Loomis v. Wisconsin über die Zulässigkeit der Verwendung des COMPAS-Algorithmus zu entscheiden.[33] Die Nutzung der Prognosesoftware wurde für zulässig befunden unter Berufung auf die Möglichkeit des Beschwerdeführers unrichtige Daten korrigieren zu können. Die der engen Sichtweise zu Art. 5 Abs. 1 lit. d DSGVO vergleichbare Ansicht des Gerichts wurde in der Literatur als den Anforderungen des Problems nicht entsprechend kritisiert.[34]

Das US-Recht bietet daher verschiedene Anknüpfungspunkte für eine Diskussion zur Datenqualität, die im Rahmen einer europäischen Debatte berücksichtigt werden können.

3.3. Technische Datenqualitätsstandards

Bei der Ausarbeitung rechtlicher Anforderungen an die Datenqualität kann auf umfassende technische Standards und entsprechende Forschungsarbeiten zurückgegriffen werden. Der Umfang und die Komplexität vorhandener Ansätze stellt die Rezeption durch die Rechtswissenschaften jedoch auch vor Schwierigkeiten.[35] Daher soll nur kurz ein Überblick zu technischen Konzeptionen der Datenqualität gegeben werden.

Die Datenqualität kann aus technischer Sicht beispielsweise als „fitness for use“ definiert werden.[36] Dieser nutzenorientierte Ansatz macht die Kontextabhängigkeit der Datenqualität deutlich. Eine andere Möglichkeit ist die Definition der Datenqualität als Maß der Übereinstimmung zwischen dem durch die Daten des Informationssystems präsentierten Bild und den tatsächlichen Daten.[37] Technische Standards sind stark ausdifferenziert mit einer Vielzahl von konkreteren Indikatoren, wie die Verlässlichkeit der Daten entsprechend der Kompetenz und Glaubwürdigkeit der datenerhebenden Stelle. Entsprechende Standards finden sich auch in technischen Normierungen, wie der ISO 8000 Norm. Aufgrund der fortschreitenden Entwicklung und der Vielzahl von Anwendungs-

---

[31] Zu den weiteren Implikationen, siehe Ga.
[32] Salt Institute v. Leavitt, 440 F.3d 156 (4th Cir. 2006), daneben haben auch andere Gruppen versucht den Data Quality Act für Deregulierungsinteressen zu instrumentalisieren: https://www.washingtonpost.com/archive/ politics/2004/08/16/data-quality-law-is-nemesis-of-regulation/3c35ae56-7935-4694-b125-348a374b657b/?utm_term=.79ca035c7e08
[33] Loomis, 881 N.W.2d at 756
[34] Zu einer umfassenden Darstellung der US-amerikanischen Diskussion hierzu, siehe Wa.
[35] Für eine Übersicht zur Forschung zur Datenqualität bei Big Data, siehe beispielsweise MBP.
[36] Siehe beispielsweise TB98.
[37] Mit weiteren Nachweisen He.

bereichen stehen allgemeingültige Standards zur Kontrolle der Datenqualität derzeit noch aus.[38]

Aufgrund der unterschiedlichen Zielsetzungen lassen sich technische Ansätze nicht ohne weiteres auf rechtliche Standards übertragen. Sie bieten jedoch einen Anknüpfungspunkt für eine fundierte und tiefergehende rechtliche Diskussion und eine Ausarbeitung differenzierter rechtlicher Standards.

4 Anwendung der Regulierung auf algorithmische Entscheidungssysteme

Fraglich ist, ob die dargestellten rechtlichen Ansätze zur Überprüfung der Datenqualität zur Regulierung algorithmischer Entscheidungssysteme geeignet sind. Dies setzt voraus, dass durch rechtliche Anforderungen mögliche Fehler vermieden werden können und dass die entsprechenden Normen praktisch anwendbar sind.

Die Rechtsdurchsetzung im Rahmen der Datenschutzgrundverordnung erfolgt zum einen durch Tätigwerden der Aufsichtsbehörden und zum anderen durch die Betroffenen selbst. Fraglich ist, inwiefern diese jeweils die Datenqualität algorithmischer Entscheidungssysteme überprüfen können.

4.1. Geltendmachung von Betroffenenrechten

Betroffene einer algorithmischen Entscheidung sind in der Regel auf Informationen des Verantwortlichen angewiesen, um deren Rechtmäßigkeit überprüfen und gegebenenfalls weitere Rechte geltend machen zu können. Angesichts der Schwierigkeiten bei der Erklärung der maßgeblichen Gründe einer algorithmischen Entscheidung[39] stellt sich die Frage, ob die Datenschutzgrundverordnung Betroffenen eine Überprüfung der Datenqualität ermöglicht.

Betroffene können gegenüber Verantwortlichen ein Auskunftsrecht nach Art. 15 Abs. 1 DSGVO geltend machen. Dies umfasst einerseits eine Auskunft über die sie betreffenden beim Verantwortlichen gespeicherten Daten und zum anderen nach Art. 15 Abs. 1 lit. h DSGVO aussagekräftige Informationen über die involvierte Logik der automatisierten Entscheidungsfindung.[40] Der Bedeutungsumfang des Begriffs der involvierten Logik ist noch nicht abschließend geklärt. Überwiegend wird hierunter eine Offenlegung der wesentlichen Entscheidungskriterien gefasst und das Verhältnis zum Ge-

[38] Siehe zu technischen Ansätzen und Standards: TSD18, CZ15.
[39] Vergleiche zu möglichen Erklärungsansätzen und deren Akzeptanz durch Betroffene: Bi18.
[40] Art. 13, 14 DSGVO normiert vergleichbare Informationspflichten des Verantwortlichen.

schäfts- und Betriebsgeheimnisschutz diskutiert.[41] Die Datenqualität wird bisher nicht erfasst. Dementsprechend umfasst der Auskunftsanspruch nach Art. 15 Abs. 1 DSGVO keine Informationen zur Datenqualität.

Betroffenen fehlt damit die Möglichkeit die Datenqualität algorithmischer Entscheidungen zu überprüfen.

4.2. Kontrolle durch Aufsichtsbehörden

Die datenschutzrechtlichen Aufsichtsbehörden können grundsätzlich die Einhaltung der Grundsätze der Datenverarbeitung nach Art. 5 DSGVO überprüfen und im Zweifel nach Art. 83 DSGVO ein Bußgeld verhängen. Es besteht daher theoretisch die Möglichkeit auf Grundlage des Grundsatzes der sachlichen Richtigkeit von Daten nach Art. 5 Abs. 1 lit. d DSGVO die Datenqualität algorithmischer Entscheidungssysteme überprüfen.

Fraglich ist jedoch die praktische Umsetzbarkeit der Kontrolle der Datenqualität. Die Aufsichtsbehörden können umfangreiche Informationen von den Verantwortlichen einfordern. Eine umfassende Überprüfung der Qualität der verwendeten Datenbestände mittels mathematischer und statistischer Verfahren würde ihre personellen und fachlichen Kapazitäten jedoch überfordern.[42] Es fehlt zum einen an der notwendigen Ausstattung der Behörden und zum anderen an dem für eine effektive Kontrolle notwendigen bereichspezifischen Knowhow. Das entsprechende Wissen, beispielsweise zum Gesundheitssektor oder der Polizeiarbeit, findet sich bei den Fachbehörden.[43] Eine technische Überprüfung der Datenqualität durch die Datenschutzbehörden wäre daher nicht zweckmäßig.[44]

In Betracht kommt daher lediglich eine Integration von Datenqualitätsstandards in den prozeduralen Ansatz der Datenschutzgrundverordnung. Maßgeblicher Anknüpfungspunkt ist hierbei die Rechenschaftspflicht gemäß Art. 5 Abs. 2 DSGVO. Der Verantwortliche ist verpflichtet, die Einhaltung datenschutzrechtlicher Anforderungen und damit auch der Grundsätze der Datenqualität im Sinne von Art. 5 Abs. 1 lit. d DSGVO nachzuweisen. Dies wird maßgeblich durch technische und organisatorische Maßnahmen, wie beispielsweise die Führung eines Verarbeitungsverzeichnisses nach Art. 30 DSGVO oder den Datenschutzbeauftragten gemäß Art. 37 DSGVO sichergestellt. Beim Einsatz algorithmischer Entscheidungen ist außerdem gemäß Art. 35 Abs. 3 lit. a DSGVO regelmäßig eine Datenschutzfolgeabschätzung nach Art. 35 DSGVO durchzuführen. Im Rahmen einer Folgenabschätzung sind mögliche Risiken der Datenverarbeitung und entsprechende Gegenmaßnahmen darzustellen. In diesem Rahmen könnte auch der Nachweis eines adäquaten Datenqualitätsmanagements als Ge-

---

[41] Sca, Di19.
[42] So Ho16a.
[43] Diese Überlegung gilt für den gesamten Bereich der Algorithmenregulierung, weswegen zum Teil eine Stärkung der sektorspezifischen Behörden gefordert wird.
[44] Auf den Konflikt zur eigentlichen Zwecksetzung der Datenschutzaufsicht hinweisend Fn. 42.

samtheit der technischen und organisatorischen Maßnahmen zur Gewährleistung der Datenqualität im Sinne von Erwägungsgrund 71 erforderlich sein. Eine Integration der Anforderungen an die Datenqualität in die Verfahrensvorgaben der Datenschutzgrundverordnung hätte den Vorteil, dass mögliche Verstöße im regulären System der Bußgeldtatbestände einfacher berücksichtigt werden können.

Für eine entsprechende Berücksichtigung von Datenqualitätsanforderungen sind verlässliche Maßstäbe erforderlich. Der Grundsatz der sachlichen Richtigkeit der Daten nach Art. 5 Abs. 1 lit. d DSGVO ist als Grundlage hierfür derzeit noch nicht hinreichend klar definiert. Eine allgemeine Pflicht zur Gewährleistung der Richtigkeit der Daten ist für die konkrete Umsetzung zu unbestimmt.[45] Zwar bieten das US-Recht sowie technische Standards Anknüpfungspunkte für eine Konkretisierung. Eine rechtssichere Festlegung von Datenqualitätsmerkmalen könnte jedoch noch einige Zeit auf sich warten lassen.

Die Datenschutzgrundverordnung bietet Ansätze zur Kontrolle algorithmischer Entscheidungen. Mangels hinreichender Begriffsklärung sind diese jedoch nicht geeignet, um die Qualität der Datengrundlage algorithmischer Entscheidungen zu gewährleisten.

5 Möglichkeiten zur Regulierung der Datenqualität

Nachdem festgestellt wurde, dass die bestehenden Möglichkeiten der Datenschutzgrundverordnung zur Kontrolle der Datenqualität algorithmischer Entscheidungen nicht ausreichen, stellt sich die Frage, wie eine Überprüfung der Datenqualität durch Betroffene und Aufsichtsbehörden anderweitig gewährleisten werden könnte.

Dies betrifft zum einen die rechtliche Definition des Begriffs der Datenqualität und zum anderen die konkrete Überprüfung. Wie bereits dargelegt, ist eine Auslegung des Art. 5 Abs. 1 lit. d DSGVO im Sinne einer klaren Trennung zwischen richtig und unrichtig für die Anwendung auf Big Data und algorithmische Entscheidungssysteme ungeeignet. Ein praktikabler Ansatz müsste sich stärker an technischen Konzeptionen von Datenqualität orientieren und verschiedene Einzelmerkmale integrieren. Dabei stellt sich die Frage, ob Art. 5 Abs. 1 lit. d DSGVO der richtige Anknüpfungspunkt ist oder ob das Wortlautverständnis durch eine entsprechende Auslegung überstrapaziert werden könnte. Alternativ könnte beispielsweise an Erwägungsgrund 71 angeknüpft werden, der sich explizit auf automatisierte Entscheidungen bezieht. Gleichzeitig würde die Etablierung divergierender Datenqualitätsstandards in der DSGVO aber nicht zur Rechtssicherheit beitragen. Daher empfiehlt sich vermutlich eine Auslegung des Art. 5 Abs. 1 lit. d DSGVO im Sinne des „accuracy"-Verständnisses.

Für die Durchsetzung mittels aufsichtsrechtlicher Kontrolle bietet sich wie bereits angedeutet eine stärkere Prozeduralisierung der Datenqualitätsanforderungen an. Dabei braucht es einerseits allgemeingültige Rahmenanforderungen sowie andererseits konkre-

[45] Siehe Fn. 25

te anwendungsspezifische Vorgaben. Insbesondere bei Letzteren wird eine Kooperation mit den zuständigen Fachbehörden notwendig sein.

Im Rahmen einer allgemeinen Richtlinie können generelle Grundsätze des Datenqualitätsmanagements rechtlich verbindlich festgeschrieben und grundsätzliche Maßstäbe für den Umfang der notwendigen Maßnahmen etabliert werden. Dabei sind insbesondere Risiken für die Betroffenen einer algorithmischen Entscheidung zu beachten. Als Grundlage solcher allgemeinen Anforderungen bietet sich die Konzeption der Datenqualität als Maß der Übereinstimmungen zwischen dem durch die Daten vermittelten Bild und der realen Situation an.[46] Dabei handelt es sich erstmal um eine bloße Zielvorgabe. Es bedürfte daher daneben einer Ausdifferenzierung durch verschiedene rechtliche Merkmale der Datenqualität. Hierbei kann zum Teil auf technische Standards oder auf Ansätze aus anderen Rechtsordnungen zurückgegriffen werden. In Betracht kommen Kategorien wie Objektivität, Vollständigkeit, Eignung für den spezifischen Kontext oder Verlässlichkeit bzw. Belastbarkeit der Daten.

Die Ausgestaltung sollte idealerweise derart erfolgen, dass der Bezug zur Gewährleistung der Datenqualität ohne weiteres nachvollziehbar ist und die Kontrolle einheitlich und zügig erfolgen kann. So könnte dem Risiko einer übermäßigen Anpassung des Algorithmus an lokale Besonderheiten durch eine Verpflichtung zum Test des Systems anhand der Daten verschiedener Standorte. So könnte beispielsweise im Gesundheitsbereich die Einbeziehung verschiedener Krankenhäuser vorgeschrieben werden. Bei einer solchen Regelung wäre zum einen der Effekt zugunsten der Datenqualität leicht ersichtlich und zum anderen die jeweiligen Testergebnisse einfach zu überprüfen. Bei der Ausgestaltung entsprechender Standards muss jedoch auch darauf geachtet werden, dass die Anforderungen an den Entwicklungsprozess nicht überfrachtet werden. Angesichts der fortschreitenden technischen Entwicklung sollten mögliche Standards flexibel bleiben, um falls notwendig angepasst werden zu können.

Für konkrete Anwendungsbereiche mit spezifischen Herausforderungen, wie beispielsweise den Gesundheitsbereich oder polizeiliche Prognosen, sind darüber hinaus explizite, den Gegebenheiten angepasste Vorgaben sinnvoll. Gegebenenfalls empfiehlt sich hier auch eine spezialgesetzliche Regelung, insbesondere beim staatlichen Einsatz algorithmischer Entscheidungssysteme wie bei der Polizei.

Entwickler und Verantwortliche algorithmischer Entscheidungssysteme müssten für den gesamten Entwicklungsprozess und während der Dauer des Einsatzes die Beachtung der Datenqualitätsanforderungen nachweisen können. Bereits als Teil der Entwicklung des Programms sind entsprechende Maßnahmen zu treffen. Der Verantwortliche ist darüber hinaus für die Lebensdauer eines algorithmischen Entscheidungssystems zur kontinuierlichen Überprüfung der Datenqualität verpflichtet. Dies wird nur in Zusammenarbeit mit dem Entwickler durchführbar sein und erfordert angemessene prozedurale Vorkehrun-

[46] Siehe hierzu Fn. 37

gen, wie beispielsweise die Implementierung von Monitoring-Schleifen.[47] Solange Unsicherheit hinsichtlich der algorithmischen Entscheidungsfindung besteht, ist es angezeigt die Entscheidungsgrundlage regelmäßig zu überprüfen und mit neuen Erkenntnissen und Entwicklungen abzugleichen.

Eine Kontrolle der Dokumentation eines entsprechenden Datenqualitätsmanagements könnte sowohl durch die Datenschutzbehörde als auch durch die jeweiligen Fachbehörden erfolgen. Jedenfalls im Rahmen einer vertieften Prüfung wird häufig eine Kooperation erforderlich sein, um auf die jeweiligen Besonderheiten angemessen eingehen zu können.

Fraglich ist, inwiefern Betroffenen eine Kontrolle der Datenqualität ermöglicht werden kann. Unabhängig von der konkreten Ausgestaltung werden hier praktische Grenzen gesetzt sein. Bereits die Erklärung des Ergebnisses gestaltet sich bei algorithmischen Entscheidungen schwierig.[48] Gleichzeitig können Überlegungen zur Vermittlung von Informationen über die Entscheidung an sich auch auf die Darstellung der Datenqualität übertragen werden. In vielen Fällen wird ein Laie nicht in der Lage sein, Fehler des Algorithmus erkennen zu können. Soweit im Rahmen der Erklärung beispielsweise Informationen über die Vergleichsdaten einbezogen werden, kann der Betroffene aber gegebenenfalls einschätzen, ob diese zu seiner Situation passen.[49] Es müsste dabei Sorge getragen werden, dass nicht lediglich die Illusion einer wirksamen Kontrolle entsteht. Grundlegende Informationen könnten jedoch Ausgangspunkt eines aufgeklärten Umgangs mit dem Problem der Datenqualität sein.

Ein solcher Informationsanspruch könnte beispielsweise durch eine weite Auslegung des Auskunftsanspruchs gemäß Art. 15 Abs. 1 lit. h DSGVO erfolgen. Als Teil der Informationen zur involvierten Logik des Entscheidungsverfahrens könnten Angaben zu den zugrundeliegenden Daten erforderlich sein. Informationen zur Datenqualität sind zum Verständnis der Entscheidungsfindung erforderlich. Eine Einbeziehung unter den Begriff der involvierten Logik wäre bei der begrifflichen Fassung der Datenbestände als Grundlage dieser Logik auch mit dem Wortlaut vereinbar.

Die bestehenden Ansätze der Datenschutzgrundverordnung können fortgeführt und weiterentwickelt werden. Die rechtliche Diskussion sollte sich dabei an der technischen Entwicklung orientieren und möglichst flexible Lösungen wählen. Für bestimmte Bereiche bietet sich eine Normierung konkreterer Vorgaben an.

---

[47] Siehe hierzu La95.
[48] Siehe zu verschiedenen Möglichkeiten hierzu Bi18.
[49] In Betracht kommt beispielsweise die Darstellung der Gesamtverteilung, wie dies beispielsweise bei Bi18 untersucht wurde. Soweit die entsprechenden lokalen Daten frei verfügbar sind, wie bei amtlichen Statistiken, könnte ein Vergleich erfolgen.

6 Zusammenfassung

Für die Kontrolle algorithmischer Entscheidungen ist es entscheidend eine angemessene Qualität der verwendeten Datenbestände sicherzustellen. Eine verzerrte oder fehlerhafte Datengrundlage führt zu fehlerhaften Entscheidungen. Die Datenschutzgrundverordnung bietet mit Art. 5 Abs. 1 lit. d DSGVO einen Anknüpfungspunkt für eine Debatte, ist jedoch für eine praktische Anwendung derzeit noch nicht hinreichend bestimmt. Ansätze aus anderen Rechtsordnungen oder technischen Standards können jedoch zur Konkretisierung herangezogen werden. Änderungsbedarf besteht auch bei den Möglichkeiten der Betroffenen sowie der Aufsichtsbehörden zur Überprüfung der Datenqualität algorithmischer Entscheidungen. Es empfiehlt sich eine Fortführung des prozeduralen Ansatzes der Datenschutzgrundverordnung. Die Information der Betroffenen kann sich an der allgemeinen Diskussion zur Erklärung von Algorithmen orientieren. Eine konsequente Überprüfung der Datenqualität stellt einen relevanten Baustein bei der Kontrolle algorithmischer Entscheidungen dar.

Literaturverzeichnis

[Ar18] Article 29 Datenschutzgruppe: Working Paper 251. Leitlinien zu automatisierten Entscheidungen im Einzelfall einschließlich Profiling für die Zwecke der Verordnung 2016/679, 2018.

[Bi18] Binns, R. et al.: 'It's Reducing a Human Being to a Percentage'. In (CHI Hrsg.): CHI 2018. Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, April 21-26, 2018, Montreal, QC, Canada. ACM, New York, NY, 2018; S. 1–14.

[CZ15] Cai, L.; Zhu, Y.: The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. In Data Science Journal, 2015, 14; S. 2.

[Di19] Dix, A.: DSGVO Art. 15: Simitis/Hornung/Spiecker, Datenschutzrecht, 2019; 25, 26.

[Ga] Gasser, U.: Information Quality and the Law. or How to Catch A Difficult Horse. In Berkman Center Research Publication, No. 2003-08.

[He] Heinrich, B. et al.: Requirements for Data Quality Metrics. In Journal of Data and Information Quality, 2018, 9; S. 1–32.

[Ho16a] Hoeren, T.: Big Data und Datenqualität - ein Blick auf die DS-GVO. Annäherungen an Qualitätsstandards und deren Harmonisierung. In ZD, 2016; S. 459–463.

[Ho16b] Hoeren, T.: Thesen zum Verhältnis von Big Data und Datenqualität. Erstes Raster zum Erstellen jurstischer Standards. In MMR, 2016; S. 8–11.

[Ki17] Kirkpatrick, K.: It's not the algorithm, it's the data. In Communications of the ACM, 2017, 60; S. 21–23.

[Kr17] Kroll, J. A. et al.: Accountable Algorithms. In University of Pennsylvania Law Review, 2017, 165; S. 633.

[La95] Ladeur, K.-H.: Das Umweltrecht der Wissensgesellschaft von der Gefahrenabwehr zum Risikomanagement, Berlin, 1995.

[MBP] Mirzaie, M.; Behkamal, B.; Paydar, S.: Big Data Quality: A systematic literature review and future research directions.

[Pa15] Pasquale, F.: The Black Box Society. The Secret Algorithms That Control Money and Information. Harvard University Press, 2015.

[Ro19] Roßnagel, A.: DSGVO Art. 5: Simitis/Hornung/Spiecker, Datenschutzrecht, 2019; 139–141.

[RSG16] Ribeiro, M. T.; Singh, S.; Guestrin, C.: "Why Should I Trust You?". In (Krishnapuram, B. Hrsg.): Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, New York, NY, 2016; S. 1135–1144.

[Sca] Schmidt-Wudy, F.: DSGVO Art. 15: BeckOK Datenschutzrecht, Wolff/Brink; 76–80.

[Scb] Schantz, P.: DSGVO Art. 5: BeckOK Datenschutzrecht, Wolff/Brink; 27–31.

[Sh18] Sheppard, B.: Warming up to Inscrutability: How technology could challenge our concept of law. In University of Toronto Law Journal, 2018, 68.

[TB98] Tayi, G. K.; Ballou, D. P.: Examining data quality. In Communications of the ACM, 1998, 41; S. 54–57.

[TSD18] Taleb, I.; Serhani, M. A.; Dssouli, R.: Big Data Quality: A Survey: 2018 IEEE International Congress on Big Data, BigData Congress 2018, San Francisco, CA, USA, July 2-7, 2018, 2018; S. 166–173.

[Wa] Washington, A. L.: How to argue with an algorithm. Lessons from the COMPAS ProPublica debate. In Colorado Technology Law Journal, 17; Issue 1.

[Wi18] Wischmeyer, T.: Regulierung intelligenter Systeme. In Archiv des öffentlichen Rechts, 2018, 143; S. 1–66.

[Wo] Worms, C.: DS-GVO Art. 16: BeckOK Datenschutzrecht, Wolff/Brink; 47–56.

Verwendung computergenerierter Kinderpornografie zu Ermittlungszwecken im Darknet[1]

Sandra Wittmer[2] und Martin Steinebach[3]

Abstract: Die Möglichkeiten für Cyberkriminelle, Straftaten mit Hilfe des Internets weitestgehend anonym zu begehen, wachsen mit der Entwicklung neuer Technologien stetig. Ein bekanntes Beispiel hierfür ist das Verbreiten von Kinderpornografie über geschützte Peer-to-Peer Netzwerke wie Tor und Freenet. Seit dem Bekanntwerden der Hintergründe des Amoklaufs vor dem Olympia-Einkaufszentrum in München ist das Darknet als Synonym für solche Netze in den Fokus der öffentlichen Wahrnehmung gerückt. Zuletzt wurde im Rahmen der 89. Konferenz der Justizministerinnen und Justizminister der Länder sogar die Zulassung von computergeneriertem kinderpornografischem Material zur Täterermittlung im Darknet angeregt. Der folgende Beitrag greift den Beschluss der Justizministerkonferenz auf und widmet sich der aktuellen Diskussion, indem auf rechtliche Rahmenbedingungen und technische Möglichkeiten zur Umsetzung eines solchen Vorhabens eingegangen wird.

Keywords: Darknet, Kinderpornografie, Strafverfolgung, Keuschheitsprobe, Computergrafik, Vektorgrafiken.

1 Einleitung

Was die Verbreitung, sowie den Erwerb und Besitz kinderpornografischer Schriften[4] i.S.d. §§ 184b ff. StGB angeht, handelt es sich um einen Kriminalitätsbereich, der in den letzten Jahrzehnten einen grundlegenden Wandel erfahren hat. Während in den Siebzigern vor allem Zeitschriften kursierten und in den Achtzigern Videokassetten ausgetauscht wurden, spielen Trägermedien mit kinderpornografischen Material heute praktisch keine Rolle mehr, da sich die Szene seit den 1990er Jahren zunehmend ins Internet – und schließlich auch ins Darknet – verlagert hat. Für Schlagzeilen sorgte zuletzt die Verurteilung der Betreiber der Kinderpornoplattform „Elysium“, die seit 2016 im Tor-Darknet existierte und zuletzt über 111.000 Mitglieder zählte. Zuvor war es den Ermittlern der Zentralstelle zur Bekämpfung der Internet- und Computerkriminalität (ZIT) in Gießen gelungen, den Serverstandort der „Elysium“-Seite aufgrund eines Programmfehlers zu lokalisieren. Anders als im Fall von „Elysium“ bleiben die Bemühungen der

[1] Das dieser Veröffentlichung zugrundeliegende Verbundprojekt „Parallelstrukturen, Aktivitätsformen und Nutzerverhalten im Darknet“ (PANDA) wurde mit Mitteln des Bundesministeriums für Bildung und Forschung unter den Förderkennzeichen 13N14355 und 13N14356 gefördert. Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor*innen.
[2] Fraunhofer SIT/TU Darmstadt, Rheinstraße 75, 64295 Darmstadt, sandra.wittmer@sit.fraunhofer.de
[3] Fraunhofer SIT, Rheinstraße 75, 64295 Darmstadt, martin.steinebach@sit.fraunhofer.de
[4] Kinderpornografischen Schriften stehen gem. § 11 Abs. 3 StGB „Ton- und Bildträger, Datenspeicher, Abbildungen und andere Darstellungen“ gleich.

doi:10.18420/inf2019_56

Strafverfolger, Serverstandorte von versteckten Diensten im Tor-Netzwerk zu ermitteln, aufgrund der im Darknet gewährleisteten technischen Anonymität von Nutzern und Diensteanbietern jedoch häufig erfolglos. Was Ermittlungen im Bereich der §§ 184b ff. StGB angeht, wird der Einsatz von verdeckt agierenden Beamten daher häufig als einzige Möglichkeit angesehen, in die pädokriminelle Szene vorzudringen. Vor diesem Hintergrund thematisiert der folgende Beitrag die Verwendung von computergenerierter Kinderpornografie zu Ermittlungszwecken im Darknet. Begonnen wird mit einer Einführung in den Phänomenbereich der sogenannten „Keuschheitsproben". Anschließend werden die rechtlichen Rahmenbedingungen für Ermittlungen im Bereich der §§ 184b ff. StGB thematisiert und auf die technischen Möglichkeiten zur Erstellung entsprechenden Materials eingegangen. Schließlich folgt ein Diskussionsteil mit Blick auf mögliche Folgeprobleme, welche die Verwendung von digitalen Missbrauchsabbildungen zu Ermittlungszwecken mit sich bringen könnte.

2 Problem: „Keuschheitsproben" auf Darknet-Plattformen mit kinderpornografischen Inhalten

Um Ermittlungsbeamten den Zugang zu Darknet-Plattformen mit kinderpornografischen Inhalten zu erschweren, wird auf einigen Seiten das Bestehen einer sogenannten „Keuschheitsprobe" gefordert.[5] Wer sich als Nutzer registrieren und am Austausch über die Plattform teilnehmen will, muss zunächst selbst strafbares kinderpornografisches Material zur Verfügung stellen. Dies kann entweder durch einen Upload der Bild- oder Videodateien in ein geschlossenes Forum oder aber durch den Versand an einen Moderator, der den User dann freischaltet, geschehen.[6] Dieser Vorgehensweise auf den Plattformen liegt die Annahme zugrunde, dass Ermittlungsbeamte zur Aufklärung von Straftaten selbst keine Straftatbestände verwirklichen dürfen. Durch das Ablegen von derartigen „Integritätsprüfungen" soll also die Strafverfolgung auf den einschlägigen Plattformen behindert werden. Um für Ermittlungen im Bereich der §§ 184b ff. StGB Abhilfe zu schaffen, wurde im Rahmen der 89. Konferenz der Justizministerinnen und Justizminister der Länder daher die Verwendung von computergeneriertem kinderpornografischem Material zur effektiven Strafverfolgung im Darknet angeregt.[7] Dies sei eine wirksame und zugleich Individualrechtsgüter schonende Methode, die eine erneute Viktimisierung der Opfer durch die Verwendung echten Materials für das Bestehen der „Keuschheitsprobe" vermeiden könnte.[8] Da in der Folgezeit vertreten wurde, dass Ermittlungsbeamte bereits nach geltendem Recht berechtigt sind, virtuelle kinderpornografische Inhalte in

[5] Fiebig, DRiZ 2019, 50 (51).
[6] Gercke, CR 2018, 480 (482, Rn. 13).
[7] Beschluss der 89. Konferenz der Justizministerinnen und Justizminister (TOP II.9), S. 2.
[8] Es gibt in jüngster Zeit Angebote von Opfern, die ihr bereits im Umlauf befindliches Material zu Ermittlungszwecken zur Verfügung stellen würden, wie Eva Kühne-Hörmann etwa in einem Interview mit dem SPIEGEL preisgab, https://www.spiegel.de/panorama/justiz/hessen-justizministerin-fordert-tabubruch-bei-kinderporno-ermittlungen-a-1211011.html. (18.06.2019)

geschlossene Foren im Darknet hochzuladen[9], sollen zunächst die rechtlichen Rahmenbedingungen für Ermittlungen im Bereich der §§ 184b ff. StGB untersucht werden.

3 Rechtliche Rahmenbedingungen für Ermittlungen im Bereich der §§ 184b ff. StGB

In Deutschland gilt – unabhängig davon, ob ein Verdeckter Ermittler i.S.v. § 110a StPO zum Einsatz kommt oder reguläre Ermittlungen i.S.d. §§ 161, 163 StPO stattfinden[10] – der Grundsatz, dass Beamte im Rahmen ihrer Ermittlungstätigkeit nicht berechtigt sind, selbst Straftaten zu begehen. Als Grund werden das Legalitätsprinzip und die andernfalls drohende Erschütterung des Vertrauens der Bevölkerung in die Integrität der Strafverfolgungsbehörden genannt.[11] In anderen Ländern – zum Beispiel in den Niederlanden – liegen die Dinge anders.[12] In der dortigen Strafprozessordnung gilt etwa das Opportunitätsprinzip, sodass von der Verfolgung von Straftaten abgesehen werden kann, wenn dies im öffentlichen Interesse ist.[13] Zudem ist es verdeckt ermittelnden Polizeibeamten in den Niederlanden unter bestimmten Umständen erlaubt, Straftaten zu begehen.[14] Obwohl diese erweiterten Ermittlungsbefugnisse bereits erfolgreich in multinationalen Ermittlungsverfahren mit deutscher Beteiligung eingesetzt wurden[15], lässt die Strafprozessrechtstradition hierzulande eine vollständige Abkehr vom Legalitätsprinzip für den Bereich der verdeckten Ermittlungen wohl nicht zu.[16] Allerdings kennt auch die deutsche Rechtsordnung Ausnahmen von Straftatbeständen für staatliche Behörden. So ist beispielsweise der Umgang mit Betäubungsmitteln nach § 4 Abs. 2 BtMG für diese erlaubnisfrei und nach § 202d Abs. 3 Nr. 1 StGB dürfen Steuer- und Strafverfolgungsbehörden zur Erfüllung ihrer Dienstpflichten mit Daten hehlen.[17] Auch für Ermittlungen im Bereich der Kinderpornografie hat der Gesetzgeber in den letzten Jahren weitgehende tatbestandliche Ausnahmeregelungen geschaffen. So wurde durch das 27. Strafrechtsänderungsgesetz schon 1993 klargestellt, dass sich Ermittlungsbeamte, die im Rahmen ihrer rechtmäßigen Pflichten handeln, weder wegen Besitzes von Kinderpornografie noch dadurch strafbar machen, dass sie einer anderen Person den Besitz verschaffen.[18] Diese gesetzliche Berechtigung zum Besitz und zur Besitzverschaffung wurde bis heute beibehalten und findet sich in § 184b Abs. 5 StGB wieder.[19] Allerdings bezieht sich die Aus-

[9] So etwa Gercke, CR 2018, 480 (484, Rn. 26).
[10] Gercke, CR 2018, 480 (482, Rn. 11.); Für Verdeckte Ermittler ist dies in RiStBV Anl.D II 2.2 ausdrücklich geregelt.
[11] Safferling, DRiZ 2019, 206 (207).
[12] Safferling, DRiZ 2019, 206 (207).
[13] Safferling, DRiZ 2019, 206 (207).
[14] Safferling, DRiZ 2019, 206 (207).
[15] So zum Beispiel bei der Übernahme und dem Weiterbetrieb des Darknet-Drogenmarktplatzes „Hansa Market“ durch die niederländischen Behörden, vgl. Safferling, DRiZ 2019, 206 (207).
[16] Safferling, DRiZ 2019, 206 (207).
[17] Safferling, DRiZ 2019, 206 (207).
[18] Gercke, CR 2018, 480 (482, Rn. 12).
[19] Gercke, CR 2018, 480 (482, Rn. 12).

nahmeregelung ausweislich ihres eindeutigen Wortlauts nicht auf § 184b Abs. 1 Nr. 1 StGB, welcher das Verbreiten (Nr. 1 Var. 1) und öffentliche Zugänglichmachen (Nr. 1 Var. 2) kinderpornografischer Schriften normiert. Dahinter steht die Überlegung, dass der Markt für entsprechende Bilder und Videos als solcher „ausgetrocknet" werden soll, um als Reflex den Missbrauch von Kindern zur Herstellung der Bilder zu bekämpfen.[20] Zwar bestehen zu Recht Zweifel daran, ob die Bereitstellung von kinderpornografischen Inhalten in rigoros abgeschirmten Foren im Darknet überhaupt ein „öffentliches Zugänglichmachen" i.S.v. § 184b Abs. 1 Nr. Var. 2 StGB darstellen kann, da es letztlich an einer Wahrnehmbarkeit der Missbrauchsabbildungen durch eine unbestimmte Anzahl an Personen fehlt.[21] Nichtsdestotrotz wird durch das Hochladen solcher Darstellungen die Tatbestandsalternative des Verbreitens aus § 184b Abs. 1 Nr. 1 Var. 1 StGB erfüllt. Auch wenn hierfür ursprünglich eine körperliche Weitergabe der strafbaren Inhalte erforderlich war, hat der BGH aufgrund der Möglichkeit, digitalisierte Daten via Internet auch „unkörperlich" weitergeben zu können, in der Zwischenzeit einen spezifischen Verbreitungsbegriff entwickelt.[22] Ein Verbreiten von Dateien im Internet liegt demnach bereits vor, „wenn die [übertragene] Datei auf dem Rechner des Internetnutzers (…) angekommen ist."[23] Dabei sei unerheblich, ob dieser die Möglichkeit des Zugriffs auf die Datei genutzt habe oder die übertragene Datei auf einem Speichermedium persistiert wird.[24] Da die Tatbestandsalternative des § 184b Abs. 1 Nr. 1 StGB sowohl real- als auch fiktivpornografische Darstellungen umfasst[25], sind computergenerierte Darstellungen vom Verbreitungsverbot ebenso betroffen wie echte Aufnahmen. Es bleibt somit festzuhalten, dass sich Ermittlungsbeamte nach geltender Rechtslage durch das Hochladen von kinderpornografischen Inhalten auf Darknet-Plattformen gem. § 184b Abs. 1 Nr. 1 Var. 1 StGB strafbar machen würden, ohne dass die Ausnahmeregelung des § 184b Abs. 5 StGB greift. Strafverfolgungsbehörden verfügen de lege lata folglich nicht über das erforderliche rechtliche Handwerkszeug, um Zugriff auf abgeschirmte Kinderporno-Tauschbörsen im Darknet zu erhalten.[26]

4 Möglichkeiten der technischen Umsetzung

Sollte de lege ferenda die Verwendung von computergeneriertem kinderpornografischem Material zur Täterermittlung im Darknet ermöglicht werden, stellt sich unweigerlich die Frage nach der Umsetzbarkeit eines solchen Vorhabens. Diese hat in den bisherigen Diskussionen um die Zulassung der „Keuschheitsprobe" jedoch keinerlei Beachtung gefunden. Aus technischer Sicht sind verschiedene Möglichkeiten zur Umsetzung denk-

[20] Safferling, DRiZ 2019, 206 (207).
[21] Gercke, CR 2018, 480 (483, Rn. 12).
[22] Palm, Kinder- und Jugendpornographie im Internet, S. 122.
[23] *BGH*St 47, 55 (58 f.).
[24] Palm, Kinder- und Jugendpornographie im Internet, S. 122.
[25] Palm, Kinder- und Jugendpornographie im Internet, S. 120.
[26] Ebenso Safferling, DRiZ 2019, 206 (207) und Krause, NJW 2018, 679 (680); zu einem anderen Ergebnis kommt Gercke, CR 2018, 480 (484, Rn. 26).

bar, die allesamt auf der Nutzung moderner Ausprägungen der Computergrafik basieren. Die grundlegende Annahme dabei ist, dass bereits heute in Computerspielen und Filmen lebensechte Nachahmungen von Geschehnissen und Personen enthalten sind, die vom Betrachter nicht oder zumindest nicht als störend wahrgenommen werden, obwohl die Inhalte in einer sehr hohen Wiedergabequalität (also mit hoher Auflösung, niedriger Kompressionsstufe und hohen Anzahl an Bildern pro Sekunde) wiedergegeben werden. Was computergenerierte „Keuschheitsproben" angeht, könnte eine niedrigere Qualität (beispielsweise durch Vortäuschen einer Videoverbindung mit niedriger Datenrate, einer Kamera minderer Qualität oder schlechter Lichtverhältnisse) sogar dazu führen, dass dem Betrachter ein Erkennen der künstlichen Natur des Bildmaterials noch schwerer fällt. Einen Beleg hierfür liefert der Fall „Sweetie" aus dem Jahr 2013, im Rahmen dessen ein computergeneriertes Kind erfolgreich als Lockvogel in Videochats mit insgesamt 20.000 Nutzern eingesetzt wurde.[27] Während der Lockvogel betrieben wurde, brachte er insgesamt 1000 Personen dazu, dem vermeintlich zehnjährigen Mädchen aus den Philippinen Geld für sexuelle Handlungen anzubieten. Technisch umgesetzt wurde „Sweetie" durch eine Kombination aus Computergrafik und dem Erfassen von Bewegungen eines menschlichen Darstellers anhand eines Motion Capture-Verfahrens, wodurch eine äußerst realitätsnahe Darstellung erzielt werden konnte. Da in der Zwischenzeit jedoch die Befürchtung geäußert wurde, dass künstlich erzeugte Bild- und Videodateien für Täter relativ leicht als Fälschungen zu erkennen seien[28], soll im folgenden Abschnitt auf die verschiedenen denkbaren Ansätze zur Erzeugung solchen Materials eingegangen werden.

4.1 Vektorgrafiken

Die einfachste Variante wäre es, analog zu Computerspielen ein Modell eines Kindes zu erstellen. Solche werden üblicherweise durch ein Drahtgittermodell in Kombination mit Texturen realisiert. Eine Unterstützung zum Erreichen eines höheren Realismus wird in Computerspielen durch Photogrammetrie[29] ermöglicht. Hier wird aus einer Vielzahl von Fotografien ein 3D Modell eines Objektes, einer Landschaft oder einer Person erstellt. Um dem Modell natürliche Bewegungen zu ermöglichen, werden menschliche Bewegungen mit Sensoren erfasst[30] oder aus Bilddaten abgeleitet.[31] Gebräuchliche Bezeichnung hierfür ist der englische Begriff Motion Capture. Im engeren Sinn handelt es sich

[27] https://www.bbc.com/news/uk-24818769, Computer-generated 'Sweetie' catches online predators, Angus Crawford, BBC News

[28] So etwa der Vorsitzende des Bunds Deutscher Kriminalbeamter Sebastian Fiedler, http://www.spiegel.de/panorama/justiz/kinderporno-ermittler-sollen-computergenerierte-bilder-nutzen-duerfen-a-1211706.html (08.04.2019).

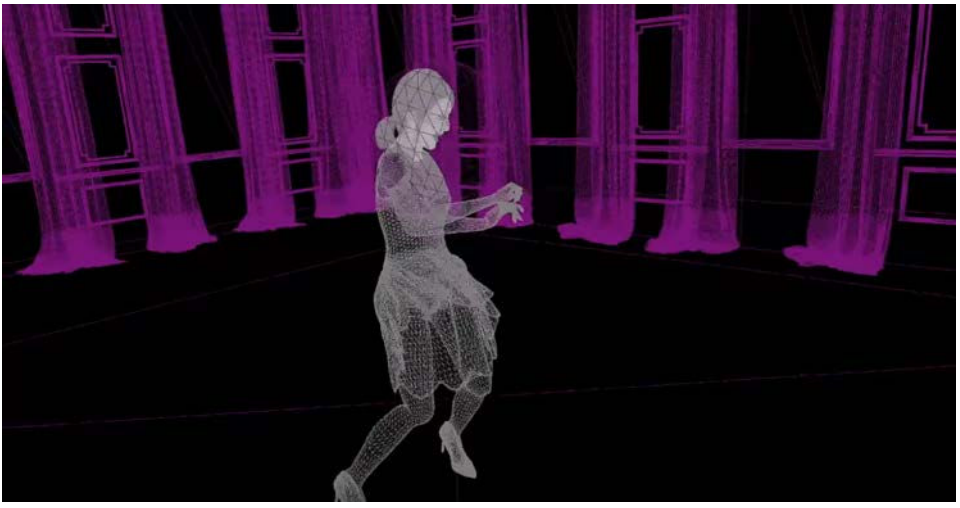
[29] Kraus, K. (2012). Photogrammetrie: Geometrische Informationen aus Photographien und Laserscanneraufnahmen. Walter de Gruyter.

[30] Vlasic, D., Adelsberger, R., Vannucci, G., Barnwell, J., Gross, M., Matusik, W., & Popović, J. (2007, August). Practical motion capture in everyday surroundings. In ACM transactions on graphics (TOG) (Vol. 26, No. 3, p. 35). Acm.

[31] Moeslund, T. B., & Granum, E. (2001). A survey of computer vision-based human motion capture. Computer vision and image understanding, 81(3), 231-268.

hierbei aber nur um das Erfassen von Bewegungen des Körpers. Die detaillierte Erfassung von Gesichtsausdrücken wird unter dem Begriff Performance Capture[32] beschrieben. Aktuelle SDKs für Computerspiele erlauben hierbei bereits fotorealistische Darstellungen von Personen, wenn die Bewegungen von Schauspielern mittels Motion Capture gesteuert werden. Die künstliche Person „Siren“, die als Demonstration für die Unreal 4 Engine eingesetzt wird, zeigt dies deutlich.

Abb. 1: Standbild der künstlichen Person „Siren“ aus dem Unreal YouTube Channel[33]



[32] Cao, C., Bradley, D., Zhou, K., & Beeler, T. (2015). Real-time high-fidelity facial performance capture. ACM Transactions on Graphics (ToG), 34(4), 46.
[33] https://www.youtube.com/watch?v=9owTAISsvwk

Abb. 2: Gitterstruktur von „Siren“, aus dem CubicMotion YouTube Channel[34]

4.2 Austausch von Gesichtern

Alternativ zu synthetisch erzeugten Inhalten könnten auch reale Bild- und Videoaufnahmen als Grundlage für das Material verwendet werden. Notwendig hierzu sind erwachsene Darsteller und Darstellerinnen, deren Körper kindlich wirken. Die Gesichter werden dann durch Methoden auf Basis maschinellen Lernens durch das von Kindern ausgetauscht. Ein bekanntes Beispiel hierfür sind die sogenannten DeepFakes, mit denen anhand eines Deep Learning-Verfahrens unter anderem die Gesichter Prominenter auf die Körper von Pornodarstellern und -darstellerinnen montiert wurden.[35] Um bei diesem Ansatz keine Gesichter von realen Kindern verwenden zu müssen, ist der Einsatz von Methoden des maschinellen Lernens denkbar[36], bei welchen Gesichter echter Kinder als Grundlage für die Erzeugung eines künstlichen Gesichts verwendet werden. Wie realitätsnah auf diese Weise künstlich erzeugte Gesichter sind, veranschaulicht die Webseite „thispersondoesnotexist“.[37] Hier werden zufällig von einem Algorithmus mittels maschinellen Lernens erzeugte Portraits gezeigt, die nur schwer von einem echten Menschen unterscheidbar sind. Erkannt werden die künstlichen Personen nur dadurch, dass der Algorithmus derzeit noch Fehler in Details macht, beispielsweise bei Haaransatz oder Augenpaaren. Fehlerhafte Bilder könnten von den Ermittlern allerdings ohne größeren Aufwand aussortiert werden.

[34] https://www.youtube.com/watch?v=zjfAPkAu2zw
[35] Harris, D. (2019). Deepfakes: False Pornography Is Here and the Law Cannot Protect You. Duke Law & Technology Review, 17(1), 99-127.
[36] Huang, R., Zhang, S., Li, T., & He, R. (2017). Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2439-2448).
[37] https://thispersondoesnotexist.com/

Abb. 3: Beispiel eines künstlichen Portraits erzeugt auf „thispersondoesnotexist.com“

4.3 Vergleich der verschiedenen Ansätze

Die oben aufgeführten Ansätze sind in der Lage, künstliches Bildmaterial zu erzeugen, welches ein Mensch zumindest bei einer flüchtigen Prüfung als „echt“ ansehen oder für eine Fotografie halten würde. Der Aufwand für die Erstellung von Videomaterial ist dabei deutlich höher als für Einzelbilder, da hier neben einer natürlich wirkenden Grafik auch noch eine ebenso natürliche Bewegung erzeugt werden muss. Bei den synthetischen Ansätzen liegt der große Vorteil darin, dass theoretisch beliebig viele künstliche Personen erzeugt werden können. Dies kann potentiell sogar sehr effizient geschehen, da das Aussehen der Personen durch einfache Parameter variiert werden kann. Im Falle von DeepFake-Ansätzen werden allerdings geeignete Körper-Double benötigt, von denen mit hoher Wahrscheinlichkeit nur eine überschaubare Anzahl existiert.

4.4 Forensische Erkennung

Aus Sicht der Multimedia-Forensik[38] sind die genannten Ansätze derzeit als erkennbar einzustufen, da synthetische Bilder von echten Fotografien unterschieden werden können. In Fotografien existieren unvermeidbar verschiedene Typen von Rauschen, die durch die Kameraelektronik verursacht werden und bei synthetisch erzeugtem Material nicht vorhanden sind. DeepFakes basieren wiederum auf Montagen, die Bilder unterschiedlicher Historie zusammenfügen und auch auf Skalierungen der Bildelemente angewiesen sind. Beide Ansätze hinterlassen also Spuren, die erkannt werden können. Die Methoden der Multimedia-Forensik stoßen allerdings schnell an ihre Grenzen, wenn ihnen mit Gegenmaßnahmen begegnet wird. Rauschen von echten Fotos lässt sich bis in hohe Details simulieren oder nachträglich von anderen Fotos übertragen[39] und Spuren, die auf Montagen hinweisen, lassen sich durch Weichzeichner oder das leichte Beschneiden des Bildes mit erneuter Kompression verwischen, da hier sowohl Hinweise auf Interpolation als auch die Historie der Kompressionsvorgänge der einzelnen Bildelemente entfernt oder stark geschwächt werden.

5 Diskussion

Obwohl die künstliche Erzeugung von virtuellen Missbrauchsabbildungen aus technischer Sicht also möglich wäre, wirft die Verwendung solchen Materials zu Ermittlungszwecken im Darknet eine Reihe von Folgeproblemen auf, welche im folgenden Abschnitt thematisiert werden sollen.

5.1 Persönlichkeitsrechte der Darsteller

Was künstlich erzeugte Bild- und Videodateien angeht, ist an die Verbindung des computergenerierten Materials zu echten Personen zu denken. Denn lediglich Einzelbilder, die aus Vektorgrafiken erstellt wurden, sind als völlig frei von personenbezogenen Daten einzustufen. Bereits die Animation solcher Vektorgrafiken in einem Video führt jedoch dazu, dass Bewegungsdaten echter Personen hinzugefügt werden. Diese Bewegungsdaten können wiederum biometrisch verwertbare Informationen beinhalten[40], die auf die echten Personen zurückgeführt werden könnten. Sobald Bilder und Videos mittels maschinellem Lernen erzeugt werden, besteht außerdem die Gefahr, dass die neu erzeugten Bilder zu nah an den Vorlagen bleiben und somit erkannt werden können. In der Praxis ist dies allerdings nur bei einer geringen Menge von Trainingsdaten zu erwarten. Das inzwischen bekannte Risiko, dass im Falle von Bildern aus trainierten Netzen die Trai-

[38] Farid, H. (2016). Photo forensics. MIT Press.

[39] Steinebach, M., El Ouariachi, M., Liu, H., & Katzenbeisser, S. (2009, September). On the reliability of cell phone camera fingerprint recognition. In International Conference on Digital Forensics and Cyber Crime (pp. 69-76). Springer, Berlin, Heidelberg.

[40] Balazia, M., & Plataniotis, K. N. (2017). Human gait recognition from motion capture data in signature poses. IET Biometrics, 6(2), 129-137.

ningsbilder wieder in hoher Qualität erstellt werden können, dürfte für das Szenario nicht kritisch sein: Ein System zum Erzeugen künstlicher Kinderpornografie wird nur in einem geschlossenen Kreis von Anwendern verbleiben dürfen, ein Verbreiten der Netze ist dementsprechend unwahrscheinlich. Im Falle von DeepFakes können jedoch auch die Körper der Darsteller und Darstellerinnen erkannt werden. Davon ausgehend, dass es sich hier um volljährige Personen handelt, die professionell Erotikdarstellungen produzieren, ist hier jedoch kein zusätzlicher Verlust an Privatheit zu befürchten.

5.2 Umsetzungsaufwand

Neben den rechtlichen Fragestellungen gilt es hinsichtlich der praktischen Umsetzbarkeit zu erörtern, ob die Erzeugung computergenerierter Kinderpornografie zu Ermittlungszwecken mit einem vertretbaren Aufwand möglich ist. Aus technischer Sicht ist dies zu bejahen – insbesondere in Anbetracht der möglichen Alternativen. Davon ausgehend, dass die „Keuschheitsprobe" nicht umgangen werden kann, ist ein erfolgreiches Bestehen nur mit Bildmaterial möglich, welches den Anforderungen der Gegenseite genügt. Da heute die Computertechnik fortgeschritten genug ist, um entsprechendes Material auf Standardrechnern zu erstellen und mit Motion Capture zu steuern, ist der Aufwand bei der Erstellung vertretbar. Von höherem Aufwand ist hier nur das Erstellen der initialen Modelle, welche gegebenenfalls auch das Mitwirken von Künstlern oder Grafikspezialisten erfordern. Da eine „Keuschheitsprobe" nach unserem Verständnis nicht in Echtzeit abgelegt werden muss, sondern auf gespeichertem Bildmaterial basiert, kann entsprechendes Material an einer zentralen Stelle produziert und den Ermittlern zur Verfügung gestellt werden. Die Anzahl von Fachleuten ist also zeitlich und räumlich begrenzt.

5.3 Befürchteter „Nachahmungseffekt"

Allerdings darf das Thema „Keuschheitsproben" nicht ohne Hinweis auf die Befürchtung diskutiert werden, dass die Verwendung von virtuellen Missbrauchsabbildungen möglicherweise den Schutzzweck des § 184b StGB konterkarieren könnte. Denn dieser beruht auf der Annahme, dass erst der durch die Verbreitung ermöglichte Konsum kinderpornografischer Materialien den Anreiz zu neuen Produktionen liefert – und damit auch den Anreiz zu immer neuem Missbrauch schafft.[41] Obwohl dieser vom Gesetzgeber unterstellte Wirkungszusammenhang bislang nicht empirisch untersucht wurde, kann nicht ausgeschlossen werden, dass der Konsum des von den Ermittlungsbeamten hochgeladenen Materials bei den Betrachtern falsche Normalität suggeriert, eigene Hemmschwellen herabsetzt und diese im schlimmsten Fall darin bestärken könnte, selbst Kinder zu missbrauchen.[42]

[41] Kuhnen, Kinderpornographie im Internet, S. 11.
[42] Kuhnen, Kinderpornographie im Internet, S. 12.

6 Fazit

Die Verwendung von computergenerierter Kinderpornografie zu Ermittlungszwecken im Darknet ist folglich eine zweischneidige Angelegenheit. Einerseits wird sich davon zwar erhofft, dem Markt für Kinderpornografie schaden zu können, indem Teilnehmer der Plattformen aus dem Verkehr gezogen werden und gleichzeitig mit dem vermehrten Auftreten von verdeckten Ermittlern in den bislang als relativ „sicher" geltenden Darknet-Tauschforen gerechnet werden muss.[43] Andererseits kann nicht ausgeschlossen werden, dass das künstlich erzeugte Material einen „Nachahmungseffekt" bei den Konsumenten hervorruft. Hinzu kommt, dass eine empirische Untersuchung, in wie vielen Fällen Ermittlungen im Bereich der §§ 184b ff. StGB de facto an einer „Keuschheitsprobe" scheitern, bislang fehlt.[44] Bevor wissenschaftlich fundierte Erkenntnisse dazu vorliegen, fällt es entsprechend schwer, sich für oder gegen die Verwendung von computergenerierten Missbrauchsabbildungen zu Ermittlungszwecken auszusprechen. Ziel dieser Ausarbeitung kann es dementsprechend nicht sein, sich diesbezüglich eindeutig zu positionieren. Vielmehr soll dieser Beitrag als Grundlage für die Diskussion eines etwaigen Reformvorhabens dienen. Erkenntnisse aus anderen Disziplinen wie beispielsweise der Kriminalpsychologie könnten in Zukunft einen wichtigen Beitrag für die abschließende Klärung der Fragestellung leisten. Feststeht, dass die Erzeugung authentisch wirkender kinderpornografischer Bild- und Videodateien aus technischer Sicht mit einem vertretbaren Aufwand möglich wäre. Dennoch sollte die erhoffte Effektivitätssteigerung der Ermittlungen gewissenhaft mit den möglichen negativen Folgen einer Ausweitung der Ermittlungsbefugnisse abgewogen werden. Wie sich das Spannungsverhältnis zwischen dem materiell-rechtlichen Verbot des § 184b StGB und der staatlichen Verpflichtung zur Aufklärung und Verfolgung von Straftaten im Falle der Einbringung einer entsprechenden Gesetzesinitiative am sinnvollsten auflösen lässt, wird letzten Endes der Gesetzgeber zu entscheiden haben.

Literaturverzeichnis

[BP17] Balazia, M., Plataniotis, K. N. (2017). Human gait recognition from motion capture data in signature poses. IET Biometrics, 6 (2), 129-137.

[Be18] Beschluss der 89. Konferenz der Justizministerinnen und Justizminister (TOP II.9), https://www.justiz.nrw.de/WebPortal_Relaunch/JM/jumiko/beschluesse/2018/Fruehjahrskonferenz_2018/II-9-BY---Effektive-Verfolgung-und-Verhinderung-von-Kinderpornografie-und-Kindesmissbrauch-im-Darknet.pdf

[Ca15] Cao, C., Bradley, D., Zhou, K., Beeler, T. (2015). Real-time high-fidelity facial performance capture. ACM Transactions on Graphics (ToG), 34(4), 46.

[43] Safferling, DRiZ 2019, 206 (207).
[44] Gercke, CR 2018, 480 (481, Rn. 7).

[Cr13] Crawford, A.: Computer-generated 'Sweetie' catches online predators, https://www.bbc.com/news/uk-24818769 (23.04.2019)

[Fa16] Farid, H. (2016). Photo forensics. MIT Press.

[Fi19] Fiebig, P.: Verbrecherjagd im Darknet, Deutsche Richterzeitung (DRiZ) 2019, S. 50-51.

[Ge18] Gercke, M.: Brauchen Ermittlungsbehörden zur Bekämpfung von Kinderpornographie im sog. „Darknet“ weitergehende Befugnisse?, Computer und Recht (CR) 2018, S. 480-484, 2018.

[Ha19] Harris, D. (2019). Deepfakes: False Pornography Is Here and the Law Cannot Protect You. Duke Law & Technology Review, 17(1), 99-127.

[Hu17] Huang, R., Zhang, S., Li, T., & He, R. (2017). Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2439-2448).

[Kr12] Kraus, K. (2012). Photogrammetrie: Geometrische Informationen aus Photographien und Laserscanneraufnahmen. Walter de Gruyter.

[Ku07] Kuhnen, K.: Kinderpornographie und Internet. Hogrefe Verlag, Göttingen, 2007.

[MG01] Moeslund, T. B., Granum, E. (2001). A survey of computer vision-based human motion capture. Computer vision and image understanding, 81(3), 231-268.

[Pa12] Palm, J.: Kinder- und Jugendpornographie im Internet. Eine materiell-rechtliche Untersuchung der Rechtslage in Deutschland. Verlag Peter Lang, Frankfurt am Main, 2012.

[Sa18] Safferling, C.: Keuschheitsproben und Verdeckte Ermittler im Darknet, Deutsche Richterzeitung (DRiZ) 2018, S. 206-207.

[Si18] Siren Real-Time Performance, https://www.youtube.com/watch?v=9owTAISsvwk (23.04.2019)

[St09] Steinebach, M., El Ouariachi, M., Liu, H., & Katzenbeisser, S. (2009, September). On the reliability of cell phone camera fingerprint recognition. In International Conference on Digital Forensics and Cyber Crime (pp. 69-76). Springer, Berlin, Heidelberg.

[Th19] This person does not exist, https://thispersondoesnotexist.com/ (23.04.2019).

[Vl07] Vlasic, D., Adelsberger, R., Vannucci, G., Barnwell, J., Gross, M., Matusik, W., & Popović, J. (2007, August). Practical motion capture in everyday surroundings. In ACM transactions on graphics (TOG) (Vol. 26, No. 3, p. 35). Acm.

Polizei und Datenschutz

Vorgaben der neuen JI-RL für technische und organisatorische Maßnahmen zur Gewährleistung datenschutzkonformer polizeilicher Datenverarbeitung

Anne Borell[1] und Stephan Schindler[2]

Abstract: Datenverarbeitung ist ein integraler Bestandteil polizeilicher Tätigkeit. Ihre Rechtmäßigkeit beschäftigt immer wieder deutsche Gerichte (zuletzt z.B. BVerfG, NJW 2019, 827 zur automatisierten Kennzeichenerkennung in Bayern). Dabei steht regelmäßig die Frage im Vordergrund, ob eine ausreichende gesetzliche Erlaubnis für die Datenverarbeitung vorliegt. Ein effektiver Schutz natürlicher Personen bei Verarbeitung personenbezogener Daten setzt aber auch technische und organisatorische Maßnahmen voraus, um sicherzustellen, dass die gesetzlichen Vorgaben eingehalten werden. Dies betrifft unter anderem die Einbindung des behördlichen Datenschutzbeauftragten, die Zusammenarbeit mit der Aufsichtsbehörde, das Führen eines Verarbeitungsverzeichnisses, die Durchführung einer Datenschutz-Folgenabschätzung, die Gewährleistung von Datensicherheit, die Beachtung von Protokollierungspflichten, die Vornahme von Maßnahmen zum Datenschutz durch Technikgestaltung und schließlich auch die Einhaltung von Benachrichtigungspflichten bei Verletzung des Schutzes personenbezogener Daten.

Keywords: Polizei, Datenschutz, Richtlinie (EU) 2016/680, technische und organisatorische Maßnahmen.

1 Einleitung

Während der Datenschutz-Grundverordnung (DSGVO) große Aufmerksamkeit zuteilgeworden ist, kann dies von der Richtlinie (EU) 2016/680 für den Datenschutz im Bereich von Polizei und Justiz[3] (JI-RL) nicht behauptet werden. Ihr Anwendungsbereich ist gemäß Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 JI-RL eröffnet, wenn die für die Verhinderung und Verfolgung von Straftaten sowie die Strafvollstreckung zuständigen Behörden zu diesen Zwecken personenbezogene Daten, also Informationen, die sich auf eine identifi-

[1] Datenschutzberaterin bei der Datenschutzberatung Moers GmbH, zuvor Wissenschaftliche Mitarbeiterin an der Universität Kassel, Fachgebiet Öffentliches Recht, IT-Recht und Umweltrecht.
[2] Universität Kassel, Fachgebiet Öffentliches Recht, IT-Recht und Umweltrecht, Henschelstraße 4, 34127 Kassel, stephan.schindler@uni-kassel.de.
[3] RL (EU) 2016/680 v. 27.4.2016 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 sowie zum freien Datenverkehr und zur Aufhebung des Rahmenbeschlusses 2008/977/JI des Rates, ABl. EU 2016 Nr. L 119/89.

doi:10.18420/inf2019_57

zierte oder identifizierbare natürliche Person beziehen (Art. 3 Nr. 1 JI-RL), verarbeiten.[4] Die DSGVO gilt in diesen Fällen nicht (Art. 2 Abs. 2 lit. d DSGVO).

Da die Verarbeitung personenbezogener Daten ein zentrales Element polizeilicher Tätigkeit zur Verhinderung und Verfolgung von Straftaten ist (z.B.: Wer hat wann was getan oder gesehen?), nimmt die Europäische Union durch die JI-RL großen Einfluss auf die innerstaatliche polizeiliche Tätigkeit. Anders als die DSGVO gilt die JI-RL allerdings nicht unmittelbar, sondern muss von den nationalen Gesetzgebern der Mitgliedstaaten erst in innerstaatliches Recht umgesetzt werden (Art. 288 AEUV). Dementsprechend sind sowohl der Bundes- als auch die Landesgesetzgeber tätig geworden. Im Folgenden wird daher exemplarisch immer auch ein Blick auf die Umsetzung der JI-RL im Bundesdatenschutzgesetz (BDSG) geworfen.[5]

Die polizeiliche Verarbeitung personenbezogener Daten geht mit Grundrechtseingriffen einher. Dies betrifft insbesondere das Recht auf informationelle Selbstbestimmung (Art. 2 Abs. 1 i.V.m. Art. 1 Abs. 1 GG[6], Art. 8 GRCh). Der rechtsstaatliche Vorbehalt des Gesetzes (Art. 20 Abs. 3 GG) sowie die grundrechtlichen Gesetzesvorbehalte fordern daher gesetzliche Ermächtigungsgrundlagen, die dem Verhältnismäßigkeits- und Bestimmtheitsgebot genügen und den mit der Datenverarbeitung einhergehenden Grundrechtseingriff auf ein angemessenes Maß beschränken. Außerdem sind die materiellen Anforderungen aus Art. 8 und 10 JI-RL[7] zu berücksichtigen.

Die damit verbundenen Fragestellungen sollen im Rahmen dieses Beitrags nicht im Vordergrund stehen. Vielmehr wird ein Blick auf die Pflicht des Verantwortlichen – also der jeweiligen datenverarbeitenden Polizeibehörde[8] (Art. 3 Nr. 8 JI-RL) – geworfen, technische und organisatorische Maßnahmen zur Sicherstellung einer datenschutzkonformen und grundrechtsschonenden Datenverarbeitung zu ergreifen. Gemeint sind Maßnahmen, die sich auf den technischen Vorgang sowie die weiteren Umstände der Verarbeitung beziehen. Dies betrifft beispielsweise die Gewährleistung von Datensicherheit (s. 2.5) oder die Einbindung eines Datenschutzbeauftragten (s. 2.1).[9]

Hinter dem Erfordernis derartiger Maßnahmen steht der Gedanke, dass effektiver Datenschutz nicht alleine durch gesetzliche Vorgaben sichergestellt werden kann. Das gilt insbesondere auch für die polizeiliche Datenverarbeitung, die nicht selten auf sensible Daten (z.B. Daten über strafrechtliche Verurteilungen) bezogen ist. Gelangen diese Da-

[4] Ausführlich zum Anwendungsbereich [HSS18].

[5] Zur Umsetzung in den Landesdatenschutzgesetzen s. Webauftritt von RA Johannes, https://lawful.de/stand-der-anpassung-der-landesdatenschutzgesetze-an-die-dsgvo-und-die-ji-richtlinie/ (Abruf 20.06.2019). Das BDSG gilt für die öffentlichen Stellen des Bundes (§ 1 Abs. 1 Satz 1 Nr. 1 BDSG), also insbesondere das Bundeskriminalamt und die Bundespolizei, soweit nicht spezifischere Vorschriften greifen (§ 1 Abs. 2 BDSG).

[6] Dazu grundlegend BVerfGE 65, 1.

[7] Es handelt sich gewissermaßen um die Gegenstücke zu Art. 6 und 9 DSGVO.

[8] Die JI-RL erfasst weitere Behörden wie Staatsanwaltschaften und Gerichte. Der Fokus liegt hier auf der Polizei als die in der Praxis maßgebliche Instanz bei der Bekämpfung von Straftaten.

[9] S. dazu [Ha18], Rn. 17. Die Aussage in der zitierten Quelle bezieht sich auf die DSGVO, ist aber auf die JI-RL übertragbar. Dies gilt im Folgenden auch für weitere Quellen.

ten in die Hände unbefugter Personen, kann dies nicht nur für die betroffenen Personen von Nachteil sein (z.B. zu gesellschaftlicher Stigmatisierung führen), sondern auch das Vertrauen in die Arbeit der Polizei erschüttern.

2 Technische und organisatorische Maßnahmen

Grundlegend fordert Art. 19 Abs. 1 JI-RL von den Verantwortlichen[10], „geeignete technische und organisatorische Maßnahmen umzusetzen, um sicherzustellen und den Nachweis dafür erbringen zu können, dass die Verarbeitung in Übereinstimmung mit dieser Richtlinie erfolgt". Dabei sind Art, Umfang, Umstände und Zwecke der Verarbeitung sowie die aus der Datenverarbeitung hervorgehenden Risiken für betroffene Personen zu berücksichtigen.

Es handelt sich um die zentrale Vorschrift zur Pflichtenstellung des Verantwortlichen, die es zumindest nahelegt, ein Datenschutz-Management einzurichten, um eine datenschutzkonforme Datenverarbeitung zu gewährleisten.[11] Als Generalklausel greift Art. 19 JI-RL, wenn speziellere Vorschriften (z.B. Art. 20 u. 29 JI-RL) nicht einschlägig sind. Vor diesem Hintergrund ist es befremdlich, dass die Vorschrift nicht in das BDSG umgesetzt wurde.

Die im Folgenden dargestellten Maßnahmen können als eine Konkretisierung der allgemeinen Pflicht aus Art. 19 JI-RL verstanden werden.

2.1 Einbindung des Datenschutzbeauftragten und Zusammenarbeit mit der Aufsichtsbehörde

Als Maßnahmen zur Gewährleistung datenschutzkonformer Datenverarbeitung sind zunächst die Einbindung des behördlichen Datenschutzbeauftragten und die Zusammenarbeit mit der zuständigen Aufsichtsbehörde zu berücksichtigen.

Einbindung des Datenschutzbeauftragten

Art. 32 Abs. 1 JI-RL (§ 5 Abs. 1 BDSG) bestimmt, dass die datenverarbeitende Polizeibehörde einen Datenschutzbeauftragten zu bestellen hat. Dessen Aufgabe besteht gemäß Art. 34 JI-RL (§ 7 Abs. 1 BDSG) insbesondere in der Überwachung der Einhaltung der datenschutzrechtlichen Vorschriften sowie der Unterrichtung und Beratung der mit der Datenverarbeitung betrauten Beamten. Dazu ist der Datenschutzbeauftragte gemäß Art. 33 Abs. 1 JI-RL (§ 6 Abs. 1 BDSG) frühzeitig in alle mit dem Schutz personenbezogener Daten zusammenhängenden Fragen einzubinden. Sollen neue datenverarbeiten-

[10] Aus Gründen besserer Lesbarkeit wird in diesem Beitrag regelmäßig davon gesprochen, dass die JI-RL von dem Verantwortlichen etwas fordert. Dies ist so zu verstehen, dass der nationale Gesetzgeber aufgerufen ist, entsprechende Regelungen zu schaffen; s. Art. 288 AEUV zur Umsetzung von Richtlinien.

[11] Dazu [Pe19a], Rn. 1 f. u. [Pe18], Rn. 1186.

de Verfahren eingerichtet oder neue Systeme angeschafft werden, ist der Datenschutzbeauftragte daher nicht vor vollendete Tatsachen zu stellen, sondern bereits in die Konzeption und die Kaufentscheidung einzubeziehen. Seine Einschätzungen sind zu berücksichtigen.[12]

Zusammenarbeit mit der Aufsichtsbehörde

Des Weiteren verpflichtet Art. 26 JI-RL (§ 68 BDSG) die Polizeibehörde zur Zusammenarbeit mit der Aufsichtsbehörde. Zuständige Aufsichtsbehörde ist, je nachdem ob es sich um eine Bundes- oder Landespolizeibehörde handelt, entweder der Bundesbeauftragte für den Datenschutz und die Informationsfreiheit (§ 9 BDSG) oder der jeweilige Landesbeauftragte. Der behördliche Datenschutzbeauftragte fungiert als Anlaufstelle für die Zusammenarbeit mit der zuständigen Aufsichtsbehörde.[13]

Während Art. 26 JI-RL die Zusammenarbeit nur auf Anfrage der Aufsichtsbehörde gebietet, enthält § 68 BDSG diese Einschränkung nicht. § 68 BDSG ist allerdings dahingehend auszulegen, dass die Polizeibehörde nicht proaktiv tätig werden muss, sondern zur Zusammenarbeit erst auf Anfrage der Aufsichtsbehörde verpflichtet ist.[14] Derartige Anfragen, etwa um einen Überblick über die gängigen Praktiken der Datenverarbeitung zu gewinnen, kann die Aufsichtsbehörde ohne Anhaltspunkte für Verstöße stellen.[15] Umgekehrt stellt sich die Frage, ob die Aufsichtsbehörde zur Beantwortung von Anfragen seitens der Polizei verpflichtet ist. Zwar kommen der Aufsichtsbehörde Beratungs- und Aufklärungsaufgaben zu.[16] Eine allgemeine Auskunftsstelle für Verantwortliche ist sie aber nicht. Dies mag unter dem Gedanken der Amtshilfe (Art. 35 GG) in Einzelfällen anders zu beurteilen sein. Unabhängig davon ist aber davon auszugehen, dass die Aufsichtsbehörde regelmäßig kooperationsbereit ist.

Die allgemeine Pflicht zur Zusammenarbeit aus Art. 26 JI-RL wird ergänzt durch spezielle Kooperationspflichten. Beispielsweise sind Verarbeitungsverzeichnisse (s. 2.2) und Protokolle (s. 2.6) auf Anforderung der Aufsichtsbehörde vorzulegen.[17] In Folge einer Datenschutz-Folgenabschätzung (s. 2.3) können sich Konsultationspflichten ergeben.[18] Hinzu treten die Untersuchungs- und Abhilfebefugnisse der Aufsichtsbehörde gemäß Art. 47 JI-RL (§ 16 BDSG). Diese umfassen unter anderem ein Recht der Aufsichtsbehörde, Zugang zu Diensträumen und Informationen zu erhalten (§ 16 Abs. 4 BDSG).

---

[12] [De19], Rn. 12 f.
[13] Art. 34 lit. d) u. e) JI-RL, § 7 Abs. 1 Nrn. 4 u. 5 BDSG.
[14] [Sc18a], Rn. 2.
[15] [Po19], Rn. 9.
[16] Art. 46 Abs. 1 JI-RL, § 14 Abs. 1 BDSG.
[17] Art. 24 Abs. 3 JI-RL, § 70 Abs. 4 BDSG sowie Art. 25 Abs. 3 JI-RL, § 76 Abs. 5 BDSG.
[18] Art. 28 JI-RL, § 69 BDSG.

2.2 Verzeichnis von Verarbeitungstätigkeiten

Art. 24 JI-RL (§ 70 BDSG) fordert, dass die Polizeibehörde ein Verzeichnis von Verarbeitungstätigkeiten (Verarbeitungsverzeichnis) führt. Dadurch soll nachgewiesen werden können, dass die Verarbeitung in Einklang mit datenschutzrechtlichen Vorgaben erfolgt.[19] Außerdem kann das Verarbeitungsverzeichnis als Mittel interner Bestandsaufnahme verstanden werden.[20] Wird es der Aufsichtsbehörde vorgelegt, erlaubt es dieser, einen Überblick über die Verarbeitung bei der Polizeibehörde zu gewinnen.

In das Verzeichnis sind gemäß Art. 24 Abs. 1 JI-RL (§ 70 Abs. 1 BDSG) unter anderem Angaben über Verarbeitungszwecke, Kategorien von Empfängern, Betroffenen und Daten sowie technische und organisatorische Maßnahmen zur Gewährleistung von Datensicherheit (s. 2.5) einzutragen. Das Verzeichnis ist regelmäßig zu aktualisieren[21] und auf Anforderung der Aufsichtsbehörde gemäß Art. 24 Abs. 3 JI-RL (§ 70 Abs. 4 BDSG) vorzulegen.

2.3 Datenschutz-Folgenabschätzung

Art. 27 Abs. 1 JI-RL (§ 67 Abs. 1 BDSG) sieht vor, dass vor Verarbeitungsvorgängen, die in Bezug auf die Rechte und Freiheiten betroffener Personen ein hohes Risiko[22], das heißt eine hohe Wahrscheinlichkeit des Eintretens eines physischen, materiellen oder immateriellen Schadens[23], aufweisen, eine Datenschutz-Folgenabschätzung durchzuführen ist. Dies gilt insbesondere bei Verwendung neuer Technologien, etwa beim Einsatz automatisierter Videoanalyseverfahren einschließlich biometrischer Gesichtserkennung. Für ähnliche Verarbeitungsvorgänge kann eine gemeinsame Folgenabschätzung vorgenommen werden (§ 67 Abs. 2 BDSG). Der Datenschutzbeauftragte ist zu beteiligen (§ 67 Abs. 3 BDSG).[24]

Der Inhalt der Datenschutz-Folgenabschätzung bestimmt sich nach Art. 27 Abs. 2 JI-RL, der in § 67 Abs. 4 BDSG umgesetzt und konkretisiert worden ist.[25] Demnach ist zunächst der Verarbeitungsvorgang zu beschreiben. Die Risiken für betroffene Personen sowie die Notwendigkeit und Verhältnismäßigkeit des Verarbeitungsvorgangs sind zu bewerten. Es müssen Maßnahmen zur Bewältigung der Risiken dargestellt werden. Das kann beispielsweise Maßnahmen zur Gewährleistung von Datenschutz durch Technikgestaltung (s. 2.4) und Datensicherheit (s. 2.5) umfassen.

---

[19] EwG 56 JI-RL.
[20] [Pe19b], Rn. 1.
[21] Dies ergibt sich zwar nicht unmittelbar aus dem Wortlaut, aber aus dem Sinn und Zweck der Vorschrift, [Pe19b], Rn. 15.
[22] § 67 Abs. 1 BDSG spricht, anders als Art. 27 Abs. 1 JI-RL, nicht von einem hohen Risiko, sondern von einer erheblichen Gefahr. Es ist aber von inhaltlicher Übereinstimmung auszugehen, [NW19], Rn. 8.
[23] Zum Schaden EwG 75 DSGVO.
[24] Als Aufgabe des Datenschutzbeauftragten Art. 34 lit. c) JI-RL, § 7 Abs. 1 Nr. 3 BDSG.
[25] Dabei lehnt sich das BDSG stark an Art. 35 Abs. 7 DSGVO an, was unproblematisch erscheint.

Ergibt die Folgenabschätzung ein hohes Risiko und werden keine Maßnahmen zur Eindämmung vorgenommen, etwa weil nach Einschätzung der Polizeibehörde dafür keine Technologien vorhanden sind,[26] oder hat die Form der Verarbeitung ein hohes Risiko für die Rechte und Freiheiten der betroffenen Personen zur Folge, ist gemäß Art. 28 Abs. 1 JI-RL (§ 69 Abs. 1 BDSG) die Aufsichtsbehörde zu konsultieren.[27] Warum diese Pflicht nur Fälle von „neu anzulegenden Dateisystemen“[28] erfasst und nicht auch für andere Verarbeitungsvorgänge im Anwendungsbereich der JI-RL gilt, ist unklar. Die Parallelvorschrift in Art. 36 Abs. 1 DSGVO kennt eine derartige Einschränkung nicht.

Im Rahmen der Konsultation sind der Aufsichtsbehörde bestimmte Unterlagen einschließlich der Datenschutz-Folgenabschätzung vorzulegen.[29] Die Aufsichtsbehörde kann Empfehlungen abgeben, wenn sie der Auffassung ist, dass gesetzliche Vorgaben verletzt werden.[30] Überdies ist die Aufsichtsbehörde nicht gehindert, ihre Untersuchungs- und Abhilfebefugnisse gemäß Art. 47 JI-RL (§ 16 BDSG) auszuüben.[31]

2.4 Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen

Stärker technisch orientiert ist die Pflicht der Polizeibehörde[32] zum Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen gemäß Art. 20 JI-RL (§ 71 BDSG).

Hinter Art. 20 JI-RL steht der Gedanke, dass effektiver Datenschutz nicht allein durch gesetzliche Vorgaben gewährleistet werden kann. Ebenso notwendig ist die Verankerung in der eingesetzten Technik. Dieser Gedanke ist freilich nicht neu, sondern lässt sich zumindest bis in die 1990er Jahre zurückverfolgen, wo er mit Namen wie *John Borking* und *Ann Cavoukian* verbunden ist.[33] Er folgt der Einsicht, dass Recht „ohne technische Unterstützung […] in einer technikgeprägten Welt folgenlos zu bleiben“ droht. Datenschutz ist dementsprechend „durch, nicht gegen Technik zu ermöglichen“.[34] Erforderlich ist daher ein Zusammenwirken von Recht und Technik, so dass bei der Entwicklung und

[26] EwG 94 DSGVO.
[27] Das Verhältnis von Art. 28 Abs. 1 lit. a) zu lit. b) JI-RL ist unklar. Insbesondere setzt lit. b) nicht voraus, dass vorher eine Folgenabschätzung durchgeführt wurde. Dementsprechend ist unabhängig von einer Folgenabschätzung und möglichen Abhilfemaßnahmen die Aufsichtsbehörde zu konsultieren, wenn die Verarbeitung ein hohes Risiko für die Rechte und Freiheiten betroffener Personen zur Folge hat, wobei in derartigen Fällen wiederum gemäß Art. 27 Abs. 1 JI-RL vorab eine Folgenabschätzung durchzuführen ist.
[28] So sowohl Art. 28 Abs. 1 JI-RL als auch § 69 Abs. 1 BDSG. Zum Begriff des Dateisystems Art. 3 Nr. 6 JI-RL.
[29] Art. 28 Abs. 4 JI-RL, § 69 Abs. 2 BDSG.
[30] Art. 28 Abs. 5 JI-RL, § 69 Abs. 3 BDSG.
[31] Darauf verweist Art. 28 Abs. 5 JI-RL, nicht aber § 69 Abs. 3 BDSG, was aber unschädlich ist, da der Aufsichtsbehörde diese Befugnisse immer zustehen.
[32] Hersteller werden durch Art. 20 JI-RL nicht adressiert. Sie sollen aber ermutigt werden, bei Entwicklung und Gestaltung ihrer Produkte den Datenschutz zu berücksichtigen, s. EwG 78 DSGVO.
[33] [BG17].
[34] [Ro05], S. 469.

Implementierung datenverarbeitender Systeme nicht nur die Wünsche der Polizei und die auf technische Machbarkeit ausgerichteten Sichtweisen der Informatiker und Ingenieure, sondern auch die Anforderungen des Datenschutzrechts von Anfang an mitzudenken sind.

Die gesetzliche Pflicht zum Datenschutz durch Technikgestaltung ist an den jeweiligen Verantwortlichen gerichtet, nicht aber an die Hersteller datenverarbeitender Systeme. Überlegungen dahingehend, auch die Hersteller zu verpflichten, wurden im Rahmen der Datenschutzreform nicht verwirklicht. Dementsprechend hat die datenverarbeitende Polizeibehörde gemäß Art. 20 Abs. 1 JI-RL (§ 71 Abs. 1 BDSG) unter Berücksichtigung von Art, Umfang, Umständen und Zweck der Verarbeitung sowie des Standes der Technik, der Kosten und der Risiken für die Rechte und Freiheiten betroffener Personen technische und organisatorische Maßnahmen zu treffen, um die Datenschutzgrundsätze (Art. 4 JI-RL) und die übrigen Anforderungen der JI-RL wirksam umzusetzen. Insbesondere sind so wenige Daten wie möglich zu verarbeiten und Daten frühestmöglich zu anonymisieren oder zu pseudonymisieren.

Derartige Vorgaben können in Konflikt mit den datengetriebenen polizeilichen Ermittlungsmethoden geraten.[35] Zu denken ist – beispielhaft – an die Videoüberwachung von Kriminalitätsschwerpunkten, gegebenenfalls in Verbindung mit Verhaltens- und Gesichtserkennung, an die automatisierte Kennzeichenerkennung oder an Predictive Policing. Es lassen sich an diesen Beispielen aber auch Möglichkeiten der datenschutzfreundlichen Technikgestaltung aufzeigen.

Bei Einsatz von Videoüberwachung ist es beispielsweise denkbar, sensible Bildbereiche (z.B. Schlafzimmerfenster) durch technische Mittel auszublenden sowie aufgenommene Personen als Pixelhaufen oder Punktwolke darzustellen.[36] Die Unkenntlichmachung kann aufgehoben werden, wenn dies aus polizeilichen Gründen notwendig erscheint. Andernfalls können die Aufnahmen nach Ablauf einer vorab festgelegten Speicherdauer automatisiert gelöscht werden. Dies kann helfen, die Grundsätze der Datenminimierung und der Speicherbegrenzung umzusetzen (Art. 4 Abs. 1 lit. c und lit. e JI-RL). Werden die Videoaufnahmen bei ihrer Anfertigung digital verschlüsselt, kann dies die Sicherheit der Datenverarbeitung stärken (Art. 4 Abs. 1 lit. f JI-RL).

Ähnliche Überlegungen können beim Einsatz automatisierter Kennzeichenerkennung greifen. In Betracht zu ziehen – und in der Praxis wohl auch weitestgehend realisiert – ist eine Gestaltung der Kennzeichenerkennung dahingehend, dass die erfassten Kennzeichen sofort mit dem hinterlegten Fahndungsbestand abgeglichen und im Nichttrefferfall unverzüglich wieder gelöscht werden, so dass sie für eine weitere Auswertung nicht zur Verfügung stehen. Ein ähnliches Vorgehen bietet sich bei bestimmten Einsatzszenarien der biometrischen Gesichtserkennung in Verbindung mit Videoüberwachung an.

[35] [Ma19], Rn. 24.
[36] [Sc19], Rn. 139.

Im Bereich des Predictive Policing können raumbezogene und personenbezogene Ansätze zur Ermittlung der Wahrscheinlichkeit des Auftretens zukünftiger Straftaten unterschieden werden. Während raumbezogene Ansätze auf Informationen zu Wetter (z.B. Sonne oder Regen) und Infrastruktur (z.B. Nachtclubs, Bushaltestellen) oder zur sozialen Zusammensetzung der örtlichen Wohnbevölkerung aufbauen, um Risikogebiete zu identifizieren, werden im Rahmen personenbezogener Ansätze Risikoprofile für einzelne Personen, etwa mit Blick auf Vorstrafen oder das soziale Umfeld, erstellt.[37] Da raumbezogene Ansätze regelmäßig ohne personenbezogene Daten auskommen, sind sie aus datenschutzrechtlicher Sicht personenbezogenen Ansätzen vorzuziehen. Andererseits können mit raumbezogenen Verfahren regelmäßig keine Aussagen über einzelne Personen getroffen werden. Ist dies gewünscht, ist die Verarbeitung personenbezogener Daten notwendig.

Gemäß Art. 20 Abs. 2 JI-RL (§ 71 Abs. 2 BDSG) ist durch Voreinstellungen sicherzustellen, dass nur die für den jeweiligen Verarbeitungszweck erforderlichen Daten verarbeitet werden. Ein typisches Beispiel dafür sind Voreinstellungen in sozialen Netzwerken, die dafür sorgen, dass bestimmte Inhalte eines Nutzers zu dessen Schutz zunächst nur für einen begrenzten Personenkreis („Freunde“) sichtbar sind.[38] Möchte der Nutzer den Personenkreis erweitern, muss er die Voreinstellungen aktiv ändern. Allerdings bieten Polizeibehörden regelmäßig keine Dienste für außenstehende Nutzer an. Die Voreinstellungen richten sich daher vor allem an Beamte, die ein datenverarbeitendes System bedienen. Dabei können insbesondere Einstellungen Bedeutung erlangen, die gemäß Art. 20 Abs. 2 Satz 3 JI-RL (§ 71 Abs. 2 Satz 3 JI-RL) dafür sorgen, dass die Daten nicht einer unbestimmten Anzahl an Personen zugänglich werden.[39] Es ist aber auch an voreingestellte Löschfristen zu denken.[40]

2.5 Sicherheit der Verarbeitung

Hinzu tritt die Pflicht der Polizeibehörde gemäß Art. 29 JI-RL (§ 64 BDSG) die Sicherheit der Verarbeitung personenbezogener Daten (Datensicherheit) zu gewährleisten.[41] Dadurch soll verhindert werden, dass Daten unbefugten Personen offenbart werden, verloren gehen oder unautorisiert verändert werden. Bei Ausgestaltung der Maßnahmen sind – wie auch in Art. 20 Abs. 1 JI-RL – Art, Umfang, Umstände und Zwecke der Verarbeitung, der Stand der Technik, die Kosten sowie die Risiken für die Rechte und Freiheiten betroffener Personen zu berücksichtigen. Es ist also ein angemessenes Maß an Sicherheit zu gewährleisten. Je höher die drohenden Schäden, desto höher die Anforderungen an die Datensicherheit.

[37] [Si18], S. 2.
[38] [Ha19a], Rn. 42.
[39] [Ma19], Rn. 35.
[40] S. Art. 20 Abs. 2 Satz 2 JI-RL, § 71 Abs. 2 Satz 2 JI-RL.
[41] Damit wird der Datenschutzgrundsatz aus Art. 4 Abs. 1 lit. f) JI-RL u. § 47 Nr. 6 BDSG konkretisiert.

Art. 29 Abs. 2 JI-RL[42] sieht einen Katalog[43] an Maßnahmen (Zugangskontrolle, Datenträgerkontrolle etc.) vor, der in regelungstechnisch merkwürdiger Weise deutlich konkretere Vorgaben als die Parallelvorschrift in Art. 32 DSGVO enthält, obwohl es sich bei der JI-RL um eine Richtlinie handelt.[44] Der zur Umsetzung geschaffene § 64 BDSG erweitert diesen Katalog noch zusätzlich (§ 64 Abs. 3 BDSG), enthält aber in Anlehnung an Art. 32 Abs. 1 DSGVO gleichzeitig die klassischen Schutzziele der Informatik – namentlich Vertraulichkeit, Integrität und Verfügbarkeit, ergänzt um Belastbarkeit und Wiederherstellbarkeit (§ 64 Abs. 2 BDSG). Überdies hat die Polizeibehörde gemäß § 64 Abs. 1 Satz 2 BDSG die einschlägigen Technischen Richtlinien[45] und Empfehlungen[46] des Bundesamtes für Informationssicherheit zu berücksichtigen.

In der Praxis kommen Maßnahmen zur Verschlüsselung von Daten, zum Logging (also zum Erstellen eines Protokolls) sowie zur Einrichtung von Rechte- und Rollenkonzepten genauso in Betracht wie bauliche Vorkehrungen (z.B. einbruchssichere Türen). Dabei ist aber zu berücksichtigen, dass zu komplexe oder zeitraubende Vorkehrungen (z.B. das Erfordernis, ständig Passwörter einzugeben) die Benutzerfreundlichkeit der Datenverarbeitungssysteme derart beeinträchtigen können, dass sie nach Möglichkeit umgangen werden (z.B. durch Weitergabe von Passwörtern oder Unterlassen des Logout).

2.6 Protokollierung

Weiterhin hat die Polizeibehörde bei automatisierten[47] Verarbeitungssystemen gemäß Art. 25 Abs. 1 JI-RL (§ 76 Abs. 1 BDSG) die Erhebung, Veränderung, Abfrage, Offenlegung einschließlich Übermittlung, Kombination und Löschung von Daten zu protokollieren. Protokolle über Abfragen und Offenlegungen von Daten müssen eine Begründung, Datum und Uhrzeit und – soweit dies möglich ist – die Identität der abfragenden oder offenlegenden sowie der empfangenden Person enthalten.[48]

Die Pflicht der Protokollierung steht im Zusammenhang mit der Führung eines Verarbeitungsverzeichnisses, da sie die Datenverarbeitung für betroffene Personen sowie die Aufsichtsbehörde überprüfbar macht.[49] Auf Anforderung stellt die Polizeibehörde die

[42] Art. 29 Abs. 2 JI-RL gilt, wie auch § 64 Abs. 3 BDSG, verpflichtend nur bei automatisierter Verarbeitung. Eine Anwendung auf die nichtautomatisierte Verarbeitung mit Bezug auf Dateisysteme, die ebenfalls in den Anwendungsbereich der JI-RL fällt (Art. 2 Abs. 2 JI-RL), wird dadurch aber nicht untersagt.
[43] Derartige Kataloge werden teilweise als antiquiert angesehen, z.B. [Er14], Rn. 48 u. 53.
[44] Darauf weist [Ha19b], Rn. 5 hin.
[45] S. Webauftritt des Bundesamtes für Sicherheit in der Informationstechnik, https://www.bsi.bund.de/DE/Publikationen/TechnischeRichtlinien/technischerichtlinien_node.html (Abruf 20.06.2019).
[46] S. Webauftritt des Bundesamtes für Sicherheit in der Informationstechnik, https://www.bsi.bund.de/DE/Themen/Cyber-Sicherheit/Empfehlungen/empfehlungen_node.html (Abruf 20.06.2019).
[47] Nicht hingegen bei nichtautomatisierten Verarbeitungssystemen. Art. 25 JI-RL ist damit enger als der Anwendungsbereich der JI-RL.
[48] Art. 25 Abs. 1 Satz 2 JI-RL, § 76 Abs. 2 BDSG.
[49] [Sc18b], Rn. 2.

Protokolldaten der Aufsichtsbehörde gemäß Art. 25 Abs. 3 JI-RL (§ 76 Abs. 5 BDSG) zur Verfügung.

Die Protokolldaten dürfen gemäß Art. 25 Abs. 2 JI-RL (§ 76 Abs. 3 BDSG) nur zur Überprüfung der Rechtmäßigkeit der Datenverarbeitung, für die Eigenüberwachung, zur Sicherstellung der Integrität und Sicherheit der personenbezogenen Daten sowie für Strafverfahren verwendet werden. Letzteres meint vor allem Strafverfahren gegen Mitarbeiter der Polizeibehörde wegen unbefugter Datenverarbeitung, etwa gemäß §§ 203, 353b StGB, die aufgrund der Protokollierung nachvollzogen werden kann.[50] Es zeigt sich darin, dass die Protokollierungspflicht auch als ein Instrument der Datensicherheit anzusehen ist, da sie spätere Kontrollen – und Sanktionen – ermöglicht und auf diese Weise durch Abschreckung dazu beitragen kann, die Vertraulichkeit und Integrität der Daten zu sichern.

2.7 Verletzung des Schutzes personenbezogener Daten

Kommt es trotz aller Sicherheitsvorkehrungen zu einer Verletzung des Schutzes personenbezogener Daten (Art. 3 Nr. 11 JI-RL, § 46 Nr. 10 BDSG), so dass Daten unbeabsichtigt oder unrechtmäßig verloren gehen oder verändert werden, oder werden Daten unbefugt offengelegt, hat die Polizeibehörde unter bestimmten Umständen die Pflicht, die zuständige Aufsichtsbehörde sowie die betroffenen Personen davon in Kenntnis zu setzen. Derartige Meldepflichten sollen die Transparenz der Datenverarbeitung erhöhen, sind Ausdruck der Rechenschaftspflicht der datenverarbeitenden Stelle und sollen gleichzeitig aufgrund des drohenden Reputationsverlustes präventiv wirken.[51]

Meldung an die Aufsichtsbehörde

Art. 30 Abs. 1 JI-RL (§ 65 Abs. 1 BDSG) fordert, dass die zuständige Polizeibehörde eine Verletzung des Schutzes personenbezogener Daten unverzüglich und möglichst binnen 72 Stunden, nachdem sie ihr bekannt geworden ist, der Aufsichtsbehörde meldet. Dies gilt nicht, wenn voraussichtlich kein Risiko[52] für die Rechte und Freiheiten natürlicher Personen, beispielsweise in Gestalt von Rufschädigung, finanziellen Verlusten oder anderen wirtschaftlichen oder gesellschaftlichen Nachteilen, besteht.[53] Vor dem Hintergrund, dass polizeiliche Datensammlungen häufig sensible Daten enthalten, etwa über Verurteilungen oder mutmaßliche Straftaten, wird ein derartiger Risikoausschluss regelmäßig nicht eingreifen.

[50] [Sc18b], Rn. 5. S. z.B. BGH, NJW 2013, 549 zur Strafbarkeit des unbefugten Zugriffs eines Polizeibeamten auf polizeiliche Datensammlung und Weitergabe an einen befreundeten Bordellbesitzer gemäß §§ 353b, 203 StGB.

[51] [Ja18], Rn. 1.

[52] Anders als Art. 30 Abs. 1 JI-RL spricht § 65 Abs. 1 BDSG nicht von Gefahr, sondern von Risiko. Inhaltliche Unterschiede gehen damit nicht einher.

[53] Beispielhaft zu drohenden Schäden EwG 61 JI-RL.

Die Meldung an die Aufsichtsbehörde muss gemäß Art. 30 Abs. 3 JI-RL (§ 65 Abs. 3 BDSG) unter anderem den Vorfall, die wahrscheinlichen Folgen und getroffene Abhilfemaßnahmen beschreiben. Überdies ist die Verletzung des Schutzes personenbezogener Daten von der Polizeibehörde gemäß Art. 30 Abs. 5 JI-RL (§ 65 Abs. 5 BDSG) zu dokumentieren. Es darf die Meldung in einem Strafverfahren allerdings grundsätzlich nicht gegen die meldepflichtige Stelle verwendet werden (§ 65 Abs. 7 i.V.m. § 42 Abs. 4 BDSG). Hierdurch soll der Nemo-tenetur-Grundsatz, also das strafverfahrensrechtliche Recht, sich nicht selbst belasten zu müssen, gewahrt werden.

Benachrichtigung betroffener Personen

Art. 31 Abs. 1 JI-RL (§ 66 Abs. 1 BDSG) verlangt eine Benachrichtigung der betroffenen Personen, also der Personen, deren Daten verändert oder offenbart worden oder verlorengegangen sind, wenn ein hohes Risiko für die Rechte und Freiheiten dieser Personen besteht. Davon ist bei polizeilichen Daten regelmäßig auszugehen. Gemäß Art. 31 Abs. 3 JI-RL (§ 66 Abs. 3 BDSG) gilt dies aber insbesondere dann nicht, wenn die Daten durch technische und organisatorische Maßnahmen, beispielsweise durch Verschlüsselung, für unbefugte Personen unzugänglich sind.

Die Benachrichtigung betroffener Personen kann überdies eingeschränkt, aufgeschoben oder ganz unterlassen werden, wenn sie die behördliche Ermittlungstätigkeit behindern würde.[54] Die Polizei ist daher beispielsweise nicht verpflichtet, einen Tatverdächtigen zu benachrichtigen, solange hierdurch ein laufendes Ermittlungsverfahren gefährdet wird. § 66 Abs. 6 BDSG bestimmt schließlich wiederum, dass die Meldung im Strafverfahren nicht gegen die meldepflichte Stelle verwendet werden darf.

2.8 Vertrauliche Meldung von Verstößen

Schlussendlich sieht Art. 48 JI-RL (§ 77 BDSG) vor, dass die zuständigen Behörden wirksame Vorkehrungen treffen müssen, um vertrauliche Meldungen über datenschutzrechtliche Verstöße zu fördern.[55] Die Möglichkeit, vertrauliche Meldungen zu tätigen, steht betroffenen Personen, Dritten sowie auch Behördenmitarbeitern offen.[56] Es handelt sich also um eine Vorschrift, die nicht zuletzt dem Schutz sogenannter Whistleblower dienen und zur Aufklärung von Missständen beitragen soll.

[54] Art. 31 Abs. 5 i.V.m. Art. 13 Abs. 3 JI-RL, § 66 Abs. 5 i.V.m. § 56 Abs. 2 BDSG.

[55] Anders als Art. 48 JI-RL sieht § 77 BDSG nicht vor, dass die öffentlichen Stellen die Meldung von Verstößen fördern. § 77 BDSG ist daher richtlinienkonform dahingehend auszulegen, dass eine Förderung stattzufinden hat, dazu [Sc18c], Rn. 2.

[56] [Sc18c], Rn. 3.

3 Fazit und Ausblick

Es hat sich gezeigt, dass die JI-RL die Vornahme verschiedener technischer und organisatorischer Maßnahmen fordert, um eine datenschutzkonforme Datenverarbeitung zu gewährleisten. Es ist allerdings nicht erkennbar, dass die – an die nationalen Gesetzgeber auf Bundes- und Landesebene gerichteten – Vorgaben der JI-RL den polizeilichen Umgang mit personenbezogenen Daten von Grund auf neu gestalten werden.

Während einige der vorgeschriebenen Maßnahmen bereits im alten Datenschutzrecht verankert waren, etwa die Pflicht zur Gewährleistung von Datensicherheit (§ 9 BDSG a.F. i.V.m. der Anlage), sind andere, etwa die Pflicht zur Führung eines Verzeichnisses von Verarbeitungstätigkeiten oder zur Durchführung einer Datenschutz-Folgenabschätzung, zwar neu, ohne aber revolutionär zu sein. In anderer Form waren sie überdies bereits im alten Datenschutzrecht angelegt.[57] Die womöglich größte Neuerung durch die JI-RL ist dann auch gar nicht in den inhaltlichen Anforderungen der Richtlinie zu sehen. Sie besteht vielmehr darin, dass mit den europarechtlichen Vorgaben, und der damit (möglicherweise) einhergehenden Bindung an die Unionsgrundrechte (Art. 51 Abs. 1 Satz 1 GRCh), der Europäische Gerichtshof als rechtsprechender Akteur hinzutritt und dem Bundesverfassungsgericht die Rolle als letztentscheidende Instanz im Bereich polizeilicher Datenverarbeitung streitig macht.[58]

Die meisten der von der JI-RL geforderten technischen und organisatorischen Maßnahmen finden sich dabei auch in der DSGVO. Dies gilt nicht für die Pflicht zur Protokollierung (Art. 25 JI-RL) und für die Förderung vertraulicher Meldungen von Verstößen (Art. 48 JI-RL). Demgegenüber enthält die JI-RL keine Vorschriften zur Zertifizierung (Art. 42 DSGVO). Dies schließt Zertifizierungen als vertrauensbildende Maßnahmen freilich nicht aus.[59] Anders als in der DSGVO sind daran aber keine unmittelbaren gesetzlichen Folgen geknüpft.[60]

Verstoßen die Polizeibehörden gegen die gesetzlichen Vorgaben, stehen der zuständigen Aufsichtsbehörde Abhilfebefugnisse zur Verfügung. Art. 47 Abs. 2 JI-RL verlangt wirksame Abhilfebefugnisse, die neben Warnungen auch das Recht umfassen sollen, Anweisungen zu erlassen und Datenverarbeitungsvorgänge zu verbieten. Die Umsetzung in § 16 Abs. 2 BDSG wird dem nicht gerecht, da die Aufsichtsbehörde die Verarbeitung nur beanstanden und eine Stellungnahme verlangen kann.[61]

---

[57] S. § 4g Abs. 2 i.V.m. § 4e BDSG a.F. zur Übersicht über bestimmte Verarbeitungsbedingungen als Vorläuferin des Verzeichnisses der Verarbeitungstätigkeiten sowie die Vorabkontrolle gemäß § 4d Abs. 5 BDSG a.F. als Vorläuferin der Datenschutz-Folgenabschätzung.

[58] [BH12], S. 152.

[59] [Ma19], Rn. 30.

[60] S. dazu Art. 25 Abs. 3 u. 32 Abs. 3 DSGVO.

[61] BT-Drs. 18/11325, S. 88 hält dies für ausreichend; kritisch z.B. [Th19], Rn. 8 f. Weitreichender z.B. die Befugnisse der hessischen Datenschutzaufsicht in § 14 Abs. 3 HDSIG.

Die Verhängung von Geldbußen gegen Behörden sieht das BDSG im Anwendungsbereich der JI-RL nicht vor (§§ 83, 84 BDSG), was im Einklang mit Art. 57 JI-RL steht.[62] Daneben bestehen gemäß § 84 i.V.m. § 42 BDSG Strafvorschriften für bestimmte Formen des unrechtmäßigen Datenumgangs.[63] Hinzu treten die Vorschriften in §§ 203, 353b StGB.

Schließlich ist in Betracht zu ziehen, dass insbesondere eine unzureichende Datensicherheit die Möglichkeit von (unbemerkten) Manipulationen eröffnet, was gegebenenfalls den Beweiswert bestimmter Erkenntnisse vor Gericht in Frage stellen und schlimmstenfalls zu Beweisverwertungsverboten führen kann.

Literaturverzeichnis

[BG17] Baumgartner, U.; Gausling, T.: Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen. Zeitschrift für Datenschutz 2017, S. 308-313.

[BH12] Bäcker, M.; Hornung, G.: EU-Richtlinie für die Datenverarbeitung bei Polizei und Justiz in Europa. Zeitschrift für Datenschutz 2012, S. 147-152.

[De19] Drewes, S.: Art. 38 Stellung des Datenschutzbeauftragten. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 890-901.

[Eh19] Ehmann, E.: § 84 Strafvorschriften. In (Gola, P.; Heckmann, D., Hrsg.): Bundesdatenschutzgesetz Kommentar, 13. Aufl. 2019, S. 768-770.

[Er14] Ernestus, W.: § 9 Technische und organisatorische Maßnahmen. In (Simitis, S., Hrsg.): Bundesdatenschutzgesetz Kommentar, 8. Aufl. 2014, S. 824-866.

[Ha19a] Hansen, M.: Art. 25 Datenschutz durch Technikgestaltung und durch datenschutzfreundliche Voreinstellungen. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 746-766.

[Ha19b] Hansen, M.: Art. 32 Sicherheit der Verarbeitung. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 813-834.

[Ha18] Hartung, J.: Art. 24 Verantwortung des für die Verarbeitung Verantwortlichen. In (Kühling, J.; Buchner, B., Hrsg.): Datenschutz-Grundverordnung/BDSG Kommentar, 2. Aufl. 2018, S. 535-543.

[HSS18] Hornung, G.; Schindler, S.; Schneider, J.: Die Europäisierung des strafverfahrensrechtlichen Datenschutzes. Zeitschrift für internationale Strafrechtsdogmatik 2018, S. 566-574.

[62] Auch im Anwendungsbereich der DSGVO hat sich der Gesetzgeber in § 43 Abs. 3 BDSG dafür entschieden, dass gegen Behörden keine Bußgelder verhängt werden, was in Einklang mit Art. 83 Abs. 7 DSGVO steht.

[63] Die Vorschriften werden teilweise kritisiert, da sie unter anderem nicht den unrechtmäßigen Umgang mit allgemein zugänglichen Daten erfassen, [Eh19], Rn. 2 und 6.

[Ja18] Jandt, S.: Art. 33 Meldung von Verletzungen des Schutzes personenbezogener Daten an die Aufsichtsbehörde. In (Kühling, J.; Buchner, B., Hrsg.): Datenschutz-Grundverordnung/BDSG Kommentar, 2. Aufl. 2018, S. 664-675.

[NW19] Nolte, N.; Werkmeister, C.: § 67 Durchführung einer Datenschutz-Folgenabschätzung. In (Gola, P.; Heckmann, D., Hrsg.): Bundesdatenschutzgesetz Kommentar, 13. Aufl. 2019, S. 670-680.

[Ma19] Marnau, N.: § 71 Datenschutz durch Technikgestaltung und datenschutzfreundliche Voreinstellungen. In (Gola, P.; Heckmann, D., Hrsg.): Bundesdatenschutzgesetz Kommentar, 13. Aufl. 2019, S. 694-703.

[Pe19a] Petri, T.: Art. 24 Verantwortung des für die Verarbeitung Verantwortlichen. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 739-745.

[Pe19b] Petri, T.: Art. 30 Verzeichnis von Verarbeitungstätigkeiten. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 802-812.

[Pe18] Petri, T.: Kap. G. XI. Verantwortlichkeit und Datenschutzorganisation. In (Bäcker, M.; Denninger, E.; Graulich, K., Hrsg.): Handbuch des Polizeirechts, 6. Aufl. 2018, S. 1097-1107.

[Po19] Polenz, S.: Art. 31 Zusammenarbeit mit der Aufsichtsbehörde. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 812-813.

[Ro05] Roßnagel, A.: Verantwortung für Datenschutz. Informatik Spektrum 2005, S. 462-473.

[Sc19] Scholz, P.: Anh. 1 zu Art. 6 Videoüberwachung. In (Simitis, S.; Hornung, G.; Spiecker genannt Döhmann, I., Hrsg.): Datenschutzrecht. DSGVO mit BDSG Kommentar, 2019, S. 463-504.

[Sc18a] Schwichtenberg, S.: § 68 Zusammenarbeit mit der oder dem Bundesbeauftragten. In (Kühling, J.; Buchner, B., Hrsg.): Datenschutz-Grundverordnung/BDSG Kommentar, 2. Aufl. 2018, S. 1569.

[Sc18b] Schwichtenberg, S.: § 76 Protokollierung. In (Kühling, J.; Buchner, B., Hrsg.): Datenschutz-Grundverordnung/BDSG Kommentar, 2. Aufl. 2018, S. 1581-1583.

[Sc18c] Schwichtenberg, S.: § 77 Vertrauliche Meldungen von Verstößen. In (Kühling, J.; Buchner, B., Hrsg.): Datenschutz-Grundverordnung/BDSG Kommentar, 2. Aufl. 2018, S. 1583-1584.

[Si18] Singelnstein, T.: Predictive Policing: Algorithmenbasierte Straftatprognosen zur vorausschauenden Kriminalintervention. Neue Zeitschrift für Strafrecht 2018, S. 1-8.

[Th19] Thiel, B.: § 16 Befugnisse. In (Gola, P.; Heckmann, D., Hrsg.): Bundesdatenschutzgesetz Kommentar, 13. Aufl. 2019, S. 178-182.

Consumer Protection in the Digital Era: The Potential of Customer-Centered LegalTech

Daniel Braun[1] Elena Scepankova[2] Patrick Holl[3] Florian Matthes[4]

Abstract: New technologies and tools, often summarised under the term "LegalTech", are changing the way in which legal professionals work. The digital transformation has changed many aspects of our daily life and democratised access to knowledge and services. In the legal domain, however, consumers rarely benefit from digitisation. On the contrary, they are often overpowered by big corporations and their well equipped legal departments. In this paper, we outline how LegalTech can be used to empower consumers in the digital era, by building tools to support consumers and those who protect them. In order to show the potential of customer-centered LegalTech, we present two prototypes which semantically analyse, assess, and summarise Terms of Services from German web shops.

Keywords: LegalTech; Customer Protection; Artificial Intelligence; Natural Language Processing

1 Introduction

The digital revolution has democratised many aspects of our lives. Access to knowledge is no longer restricted to those who can afford 32 volumes of Encyclopædia Britannica or have access to university libraries, instead it is available to everyone with access to the internet. The access to the fine arts, but also to once expensive services like translation, was opened up to new classes of citizens by digitisation.

For a long time, the legal domain was arguably one of the biggest resistance to digitisation efforts and in some aspects still struggles to catch up with other industries. A fact painfully displayed by the case of the "special electronic attorney mailbox" ("besonderes elektronisches Anwaltspostfach", beA) [MH18]. Nowadays, digitisation has entered the legal profession as so-called "LegalTech", a portmanteau word consisting of "legal services" and "technology", widely used as description for the support or automation of legal processes with software or online services.

[1] Technical University of Munich, Department of Informatics, Boltzmannstraße 3, 85748 Garching, Germany daniel.braun@tum.de
[2] Technical University of Munich, Department of Informatics, Boltzmannstraße 3, 85748 Garching, Germany elena.scepankova@tum.de
[3] Technical University of Munich, Department of Informatics, Boltzmannstraße 3, 85748 Garching, Germany patrick.holl@tum.de
[4] Technical University of Munich, Department of Informatics, Boltzmannstraße 3, 85748 Garching, Germany matthes@tum.de

doi:10.18420/inf2019_58

However, unlike in other areas, mostly big companies and law firms benefit from these developments. Almost all of the existing LegalTech tools, like Lexis Advance[5], rfrnz[6], Juristische Textanalyse[7], and Lawlift [8], to name just a few, are made for companies and law firms, rather than consumers. Therefore, LegalTech tools are not only missing the opportunity to democratise the access to legal advice by making it more affordable and available, they are actively supporting the current imbalance of power, existing between companies and consumers[9], by providing companies with even more advantages over customers.

Currently, there are only a few examples of LegalTech tools, like Flightright[10] or Chevalier[11], which are build for the benefits of consumers. And even these tools are still build to serve the commercial interests of their operators. In this paper, we want to advocate the idea of customer-cenetred LegalTech tools and present two prototypes which semantically analyse, assess, and summarise Terms of Services from German web shops.

2 Significance of Terms of Services

Standard form contracts trace back to the 19th century. In the age of industrialisation, entering into contracts has been accompanied by the unilateral use of pre-formulated rules tailored to one party's own interests and thus resulting in an imbalance of powers between the contracting parties. [Ze14] Today, customers are confronted with standard form contracts every day in form of Terms of Services (ToS), for example when they buy something online or register for an online service. Studies with more than 45,000 participants have shown that only 0.1% to 0.2% of customers read ToS of online shops. [BMWT14]

While standard form contracts regularly reflect an imbalance of contracting power, this imbalance is even stronger in situations where one of the contracting parties is a consumer without a professional legal background and the other one is a company with a potentially huge legal department. The relevance of this is visible by the amount of jurisprudence in this area, comprising more than 28,000 judgements in Germany only. [JU17]

In acknowledgement of these facts, the European lawmaker has limited the creative leeway for companies, when it comes to standard form consumer contracts.[12] One might ask oneself why consumers should care about unlawful clauses in ToS, because in case of a dispute,

[5] https://www.lexisnexis.com/en-us/products/lexis-advance.page
[6] https://rfrnz.com/
[7] https://www.datev.de/web/de/top-themen/rechtsanwaelte/juristische-textanalyse/
[8] https://www.lawlift.de
[9] Cf. German Constitutional Court BVerfGE 89, 214 of 19 October 1993.
[10] https://www.flightright.de/
[11] https://www.chevalier.law/
[12] Council Directive 93/13/EEC of 5 April 1993 on unfair terms in consumer contracts, OJ L 95/29 of 21 April 1993.

these clauses are void anyway. In reality, however, at least for online shopping, the amount in dispute is often so low that consumers avoid legal steps, even if they are in the right.

3 Related Work

A goal similar to the one outline in this paper is pursued by the project «Terms of Service; Didn't Read» (ToS;DR). [BM14] Instead of automatically assessing and evaluating the ToS, ToS;DR is crowd-sourced and provides manually generated summarisations of ToS from many major websites. However, the fact that ToS;DR is crowd-sourced affects their scalability and topicality. An automated approach to analyse online standard form contracts was presented by Lippi et al. Their analysis focuses solely on so-called «unfair clauses» which are forbidden under the law of the European Union. [Li17] In their experiment, they analysed Terms of Services from 20 major websites regarding eight unfair clauses. In a leave-one-document-out evaluation, they achieved a precision of 0.62 using a Support Vector Machine. In contrast to our approach, Lippi et al. try to do a binary classification (unfair clause exists / does not exist), while we try to gather additional information and summarise them. In order to create these summaries, a system first has to obtain the relevant information from the text. Information Retrieval (IR) for legal texts has gained a lot of attraction in recent years. Examples are McCallum [Mc05], Grabmaier et al. [Gr15], Francesconi et al. [Fr10], and Shulayeva et al. [SSW17], or, for German texts, Walter and Pinkal [WP06], and Watlt et al. [Wa17]. The issue of simplifying legal texts was addressed by Bhatia et al. [Bh83], Collantes et al. [Co15] and others.

4 Possible Approaches

We identified two possible approaches to tackle the imbalance of power between consumers and web shop operators:

1. **Directly support consumers**: By automatically finding, assessing, and summarising ToS with regard to lawfulness and customer-friendliness, we can empower consumers to make educated decisions about where to buy or not. We first presented this idea in [Br17].
2. **Support consumer protection agencies**: Instead of directly supporting consumers, we can also support those organisations who protect consumers by providing them tools to automatically analyse large amounts of ToS. Unlike consumer themselves, such organisations are often willing to take on legal battles with companies about their ToS and can therefore support the enforcement of existing customer protection laws.

The consumer protection law aims to protect the consumer and allow for «optimal market decisions» [Oe06]. The aim to enable consumers to take «optimal market decisions» is

considered as fundamental and is intended to foster the faith of the consumers into market processes. [Ta11, p. 19] German consumer law uses different regulatory instruments to achieve this, by preventively designing consumer protection law and by subsequently declaring contractual clauses not in accordance with those legal provisions as «void». In order to work effectively, legal regulations comprise both market-complementary as well as market-compensatory instruments [Re08, p. 47].

With the first approach, we likewise aim to enable the consumer to take «optimal market decisions». We consider both above-mentioned ways as important, but we think that there are obstacles which hinder legal regulations to achieve full value. The most important fact is that reading and understanding legal contracts is hindered by the fact that in comparison to regular language, legal language is characterised by a high degree of abstractness. In order to fulfil its function as a merit instance for any socially relevant behaviour, law itself needs to guard its capability of abstractly reacting even to unforeseen situations, which results in formulations characterised by a low level of comprehensibility. By summarising contractual terms and translating their legal and linguistic complexity into a simplified language (summary generation), we provide the customer with the possibility to understand his rights and duties and to take decisions based on knowledge, not on – justified or unjustified – trust towards the shop provider.

Secondly, standard form contracts like Terms of Services address fundamental conditions of performance of a certain business. In the context of online shopping, they set the provisions for e.g. payment, delivery, revocation or liability. Due to the abstract character of the legal language stipulating contractual rights and duties on the one hand, and normative requirements in legal regulations and judicial decisions on the other hand, the consumer often finds himself not capable to understand and assess the validity or invalidity of his/her contract. With our second functionality, we thus automatically (a) identify the differences between the clauses of (online shops as) companies and (b) assess their (un)lawfulness.

While this approach directly supports consumers, it has one shortcoming: It does not help to make ToS fairer and only supports those who use the proposed tool. In order to get rid of illegal ToS clauses for good, legal actions are necessary in order to force companies to change their ToS. In Germany, the “Verbraucherzentralen” (customer protection agencies) are important actors when it comes to the legal enforcement of consumer interests. They regularly admonish web shop provider for illegal ToS clauses. Therefore, we teamed up with experts from the Verbraucherzentralen to develop them in their daily work. Unlike consumers, they are interested in checking large amounts of web shops in a short time and also in re-checking them after a while. Moreover, they often are interested in more complicated legal issues than consumers. A tool targeted at this group of professional users should take this into account.

5 Technology

As both approaches have different requirements regarding the depth of the legal analysis that needs to be performed, we decided to also implement them using different technologies. However, they still have very much in common: We use a pipes and filters architecture for both prototypes and the first steps of the pipeline are identical for both prototypes.

5.1 ToS Page Classification

First, we use a naive Bayes classifier to find the ToS page of a web shop by classifying each linked sub-page as "ToS" or "Other". The classifier was trained with a set of 400 manually annotated pages from web shops, 200 of them ToS pages and 200 of them other pages. While the classifier performs very well (cf. Section 7.1 for an evaluation), it is not very fast. In order to be able to present results to consumers as fast as possible, we decided to adopt a hybrid approach and developed a rule-based URL classifier, to pre-select sub-pages that potentially are ToS page and hence restrict the set of pages that have to be classified by the naive Bayes classifier.

The classification is realised using a rule-based approach that matches common patterns for ToS links. One common pattern we identified is that the URL often contains *"AGB"*. The classifier separates URL strings into the following components: scheme specifier, network location part, path, query parameters. The path and query parameters are matched against a set of pre-defined, weighted rules. If a candidate reaches a certain threshold, we consider that a given URL points to a potential ToS page.

5.2 Content Extraction

The page which was classified as ToS is then further processed to extract the actual content from the website by e.g. removing headers, navigation etc. The current prototypes use the open source Java-library boilerpipe[13] for this task.

5.3 Information Extraction

After this step, the pipelines for both prototypes differ. For the consumer-oriented prototype, we use simple POS-tagging, for the professional-oriented prototype we use the neural network from the StanfordCore NLP[14] library to build a dependency tree for each sentence of the ToS. Afterwards, for both prototypes, the annotated ToS are analysed by a rule-based information extraction tools, which extracts the important information from the text and stores it in a JSON-format. The format of the extracted information is shown in Listing 1.

[13] `https://boilerpipe-web.appspot.com/`
[14] `https://stanfordnlp.github.io/CoreNLP/`

```
{
    "topic": "Widerrufsrecht",
    "dataType": "Timespan",
    "value": 30,
    "unit": "Tag"
}
```

List. 1: Format of extracted information

The information extraction is a two-step process. In the first step, we identify for each sentence, whether it contains a topic of interest, like information about the return policy or limitations of the form of termination. For this step, we mainly use a keyword search which works on a stemmed version of the ToS. Due to the highly regulated nature of legal language this rather simple approach is still very effective. For example, in order to legally restrict the possibility to send back goods in Germany, the term “Widerruf” has to be used.

Once the relevant sentences are identified and labelled with their topics, the second step is to actually extract the information contained in the sentence. In order to do this, a set of rules is stored for each topic. Currently, these rules have to implemented in the source code, in the future we would like to provide users with a mean to create rules during run-time, e.g. with a graphical interface or a domain specific language.

For the consumer prototype, we use regular expressions for information extraction. Table 1 show examples of different extractions rules, translated from German.

	unlawful	**rules (translated from German)**
Right of warranty	New goods: less than 2 years; used goods: less than 1 year	-warranty . . . ([0-9]* | [one | two | . . .]) [day(s) | month(s) | year(s)] AND used OR NOT used (goods | products) -warranty . . . used (goods | products) . . . excluded
Right of withdrawal	Products have to be send back using the original packaging	-product . . . original (packaging | packed | . . .) . . . (return | send back) -original (packaging | packed | . . .) . . . (return | send back)
Period for withdrawal	Period of less than 14 days for shops trading in the EU	-withdraw . . . ([0-9]* | [one | two | . . .]) [day(s) | month(s) | year(s)]
Obligation to inspect product	Warranty rights only if customer inspects and/or reports any product defects	-warranty . . . [inspect|report] AND NOT merchant
Risk of loss	In case of shipped sales the customer bears the risk of loss	[risk of loss | bearing the risk] . . . [shipped | carriage of goods] . . . consumer

Tab. 1: Extraction rules from the consumer prototype

For the professional prototype, we first generate a dependency tree for each relevant sentence. The nuances that come with more complex legal issues are often difficult to tackle with regular expressions. The limitation of the form of termination is a good example of this. Very

often, one can find in the same sentence forms of termination being listed as permissible and others being ruled out. The sentence "Die Kündigung muss schriftlich oder telefonisch erfolgen, eine Kündigung per E-Mail ist nicht möglich." (The termination must be in written form or through phone, a termination via email is not possible.) specifies that terminations have to be written or made through phone and at the same time rules out terminations via email. With regular expressions, it is difficult to write rules to cover all possible permutations of such a sentence, especially in German, and figure out which parts are included in the negation.

Dependency trees, such as the one shown in Figure 1, make the dependencies between words explicit and can therefore help to analyse which part of a sentence is negated and many other things. Instead of writing rules which just analyse the words, with dependency trees, we can also analyse the structure of the complete sentence and hence conduct more fine-grained analyses in order to extract structured information from the text. [Br18]

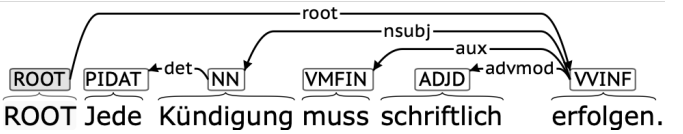


Fig. 1: Dependency Tree for the Sentence "Jede Kündigung muss schriftlich erfolgen." (Any termination must be in writing.)

5.4 Assessment

The structured representation of the information contained in the ToS is used for the legal assessment. This assessment is based on a knowledge-base which contains information about legal regulations, like the right to return or warranty rights. The database contains for example the information that, in the European Union, customers have the right to return a product they bought online for at least 14 days.[15] For the consumer version, we use three labels: "legal", "illegal", and "customer friendly". For the given example of return policies, a time span of fewer than 14 days would be classified illegal, 14 days as legal, and everything above as customer friendly. For the professional prototype, we just use the labels "legal" and "illegal". Currently, the knowledge-base is stored in a simple JSON-format and has to be directly managed as a file. In the future, we would like to build a graphical user interface which allows domain experts to maintain and extend the knowledge-base more easily.

5.5 Summarisation

In order to summarise the extracted facts in a simplified language, we currently use a simple template-based approach. Examples of the generated summarisations can be found

[15] While there are different exceptions from this rule, e.g. for individualised goods, the current prototype does not take into account these special cases.

in Figures 3 and 5. In the future, we would like to make the text generation more flexible by replacing the templates with a surface realiser like SimpleNLG [GR09, Bo11].

6 Prototypes

We developed two prototypes in order to show the potential of LegalTech applications to empower consumers and those who protect them. The first prototype, dubbed "consumer prototype" (cf. Section 6.1), was developed in order to serve people without any specialist legal knowledge while shopping online. The second prototype ("professional prototype", cf. Section 6.2) was built with experts from the customer protection agencies in mind as users who want to analyse large amounts of ToS at once. Both prototypes are implemented as web applications, so users do not need to locally install any software. Both backends were implemented using Java and a REST-API to communicate with the frontend.

6.1 Consumer Prototype

In the consumer prototype, the URL of a web shop can be entered through an input field. After clicking on the "Find ToS" button (cf. Figure 2), the backend classifies all outgoing links of the input page with the classifier described in Section 5.1. A breadth-first search is conducted until a ToS page was identified. As an intermediate result, the URL of the ToS page is shown next to the button.

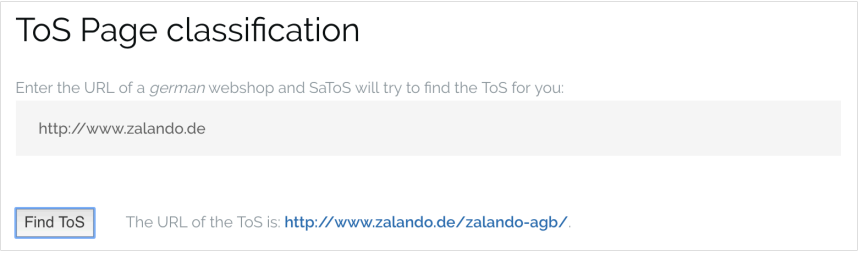


Fig. 2: Input mask of the consumer prototype

Once the URL was identified, the content extraction, information extraction, assessment and summarization are conducted. The results of all these steps are presented as shown in Figure 3. The results are split into two topics: revocation & right to return ("Widerruf & Rückgaberecht") and warranty ("Gewährleistung"). On the left side of the table, the original texts are shown which are excerpted from the ToS. Colour-highlightings are used to explain based on which parts of the text the assessment algorithm made its decision. Green highlights show identified time limits and blue highlights show identified topics. In the middle of the table, the assessment can be found. A "thumbs up" means a clause was identified to be customer-friendly, i.e. exceeds the legal minimum in a way which is beneficial for consumers. A "thumbs down" means a clause was identified as illegal and

a “neutral position” means that a clause fulfils the legal minimum. On the right side, the automatically generated summarisation is shown.

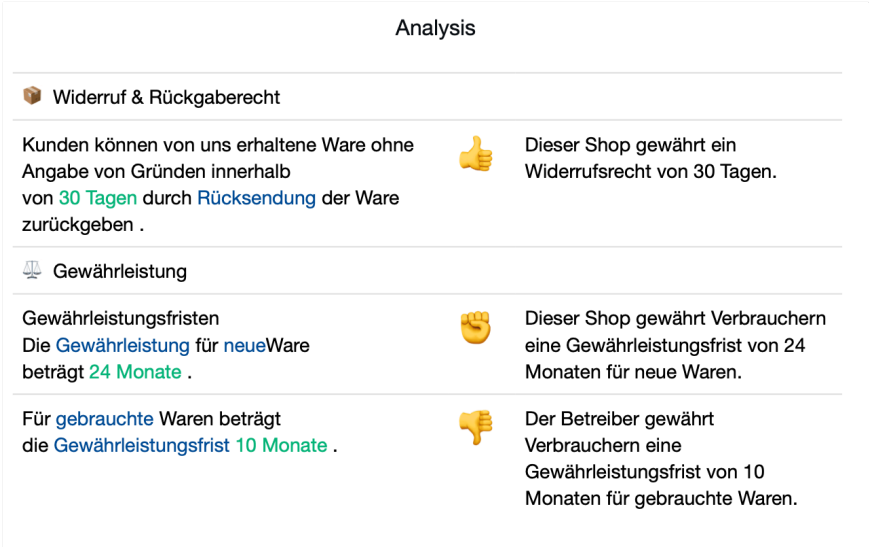


Fig. 3: Consumer prototype (mock ToS have been used as input data)

While the presented prototype is currently just a proof of concept, running such a service “in production”, i.e. giving consumers direct access to it, does have legal implications. First, liability is a question, but second, there is the German Act of Out-of-Court Legal Services which stipulates that in individual cases legal advice against payment shall only be provided by legal personnel, e.g. lawyers, legal counsels, and tax consultants.[16] One could argue that our functionality serves as a mere clarification tool operated on standard form contracts in a general way. As we neither intend nor are capable of providing specific legal advice in individual cases, our tool focuses on enhancing the understanding of legal language. However, these are open questions which go beyond technical feasibility, which have to be solved in order to allow consumers to benefit from LegalTech.

Moreover, it needs to be clearly communicated to potential users, which aspects of the ToS are covered by the analysis and that the tool just rates these specific aspects and does not make any statements about other aspects of the ToS, nor does it provide an assessment of the ToS as a whole.

[16] Cf. § 8 Abs. 1 Nr. 4 Act of Out-of-Court Legal Services.

6.2 Professional Prototype

The prototype which we developed to support domain experts from the customer protection agencies has a different interface which is tailored towards their own needs. As shown in Figure 4, the interface has three possible input types. On the top, there is an input field where the user can enter a URL. In contrast to the consumer prototype, it is also possible to enter multiple URLs, separated by commas. In addition, it is possible to upload a PDF file or directly paste plain text into the input box.

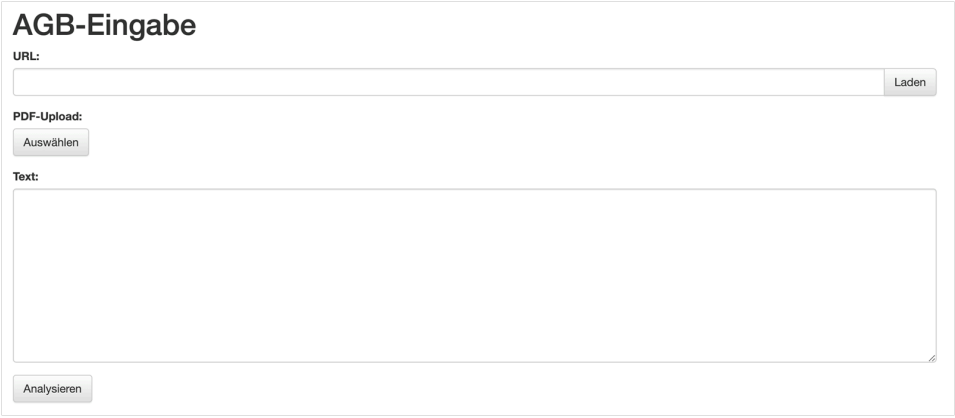


Fig. 4: Input mask of the professional prototype

In the first step, all inputs are converted to plain text. In case of a URL this means that the same steps are performed as in the consumer prototype: identification of the ToS page and content extraction. In case of a PDF file as input, the text is extracted using Apache PDFBox[17]. Afterwards, information extraction, assessment, and summarization are performed and the results are displayed as shown in Figure 5.

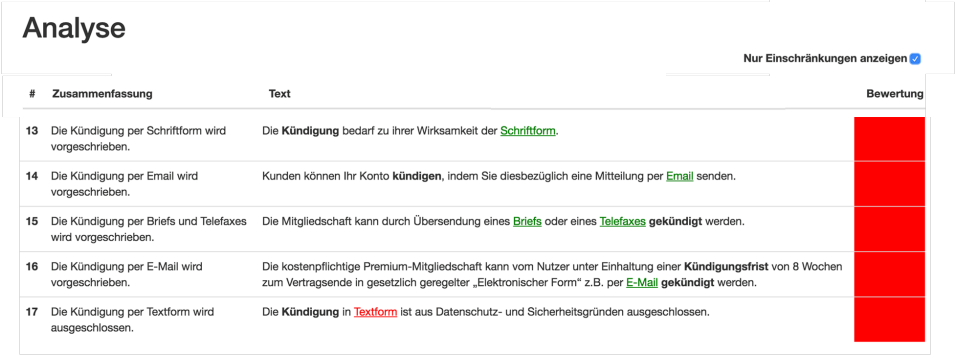


Fig. 5: Professional prototype (mock ToS have been used as input data)

[17] https://pdfbox.apache.org/

Other than the consumer prototype, which is optimised to show the analysis of one ToS page, the professional prototype is optimised to show large amounts of data. It therefore includes a filtering function which allows to only show clauses which were identified as illegal ("Nur Einschränkungen anzeigen"). Other than that it contains the same three elements: the original text with highlighting to make the reasoning transparent (this time in the middle), the assessment on the right, which is only binary in this prototype (red = illegal, green = legal), and a short summary on the left.

7 Evaluation

Since the professional prototype uses more advanced technology, the evaluation will focus on this prototype. Moreover, since all further analyses are based on the correct classification of ToS pages, we conducted a separate evaluation for the ToS page classification.

7.1 ToS Page Classification

As mentioned in Section 5.1, the prototypes use a hybrid approach of rule-based and machine learning classification for ToS pages. We collected a dataset of 3424 pages from web shops. 2592 ToS pages, manually labelled by a price comparison website, and 832 other web shop pages. We split the dataset into training (200 ToS and 200 Other) and test (2392 ToS and 632 Other) data.

The results of the evaluation are shown in Table 2. It is visible that the ML approach performed significantly better. Given the relatively small dataset which was used to train the classifier, these results are very promising. The fifth column in Table 2 shows the average time in seconds, that was needed to classify a URL. If successful, the rule-based approach is, as expected, much faster.

approach	precision	recall	F-score	$\varnothing t$ in s
ML	0.9115	0.8219	0.8644	1.435
rule-based	0.7953	0.5393	0.6428	0.001

Tab. 2: Evaluation ToS Classification

7.2 Classification of limitations of the termination form

In order to evaluate the classification of limitations of the termination form, which is used by the professional prototype, we used a hand-picked sample of 25 ToS pages from online subscription services, like fitness platforms. In each ToS page, sentences which contain limitations of the termination form were manually labelled with either legal or illegal. In

total, the 25 ToS contain 23 sentences with limitations of the termination form of which six contain illegal limitations.

Our algorithm detected all 23 sentences correctly that contained limitations (true positives) but falsely identified five more sentences (false positives). On this test data set, our algorithm achieved therefore a recall of 1.0, a precision of 0.81 and an F-score of 0.9 when it comes to the detection of sentences which contain limitations of the termination form. When it comes to classifying whether these limitations are illegal or not, the algorithm correctly labelled all six illegal limitations (true positives). One legal limitation was falsely classified as illegal (false positive). This means a recall of 1.0, a precision of 0.85, and an F-score of 0.92.

While the data set might be too small to draw general conclusions, the results are very promising and confirm the assumption that dependency trees might be a viable technology for the problem at hand. The algorithm was designed in a way which is optimised towards a higher recall which is also reflected by the results of the evaluation. Since the system is designed as a mean of support for human experts, false positives can be quickly identified by the humans in the loop. False negatives, on the other side, may never be noticed because the amount of data is too large to manually check every sentence for limitations of the termination form.

In the future, it would be desirable to investigate how experts and consumers react to the respective prototypes, in order to find out whether they could really have a real-world impact.

8 Vulnerabilities

The evaluation above measures the performance of the proposed approach in a controlled environment. However, assuming one of the described systems would be successfully applied in the real world, there are other aspects which would have to be taken into account. Most importantly, companies could try to cheat the system by optimising their ToS to disguise their consumer unfriendly clauses from the algorithm. A similar situation can be seen when it comes to optimising websites for search engine algorithms. [Ma08] This is a possible vulnerability of the consumer prototype. For the professional prototype, we expect organisations to not share their set of rules by which they evaluate ToS. This makes it difficult for companies to find out based on which formulation an assessment was made.

In general, there is no easy solution to fix this vulnerability. In the area of search engine optimisation, we have seen an arms race between search engines and dubious website providers for years. Consequently, it would also be necessary in the case of ToS assessment to regularly update the rules to detect the latest fraud attempts of unfaithful shop providers. This also emphasises the importance of customer protection agencies, which could help to hold such providers legally accountable for their deception.

9 Conclusion

In this paper, we present two approaches to consumer protection in the digital era by empowering consumers and those who protect them with LegalTech. We chose ToS as the subject of our research because consumers are confronted with them multiple times every day. Beyond presenting two prototypical implementations, which semantically analyse and asses ToS from German web shops, our goal is to raise awareness for the fact that, as of today, mostly companies benefit from the digitisation in the legal industry. This cements the already existing imbalance of power between consumer and companies. Moreover, we want to spark a discussion about the legal issues which have to be resolved, before consumers can truly benefit from LegalTech applications (namely the questions of liability and out-of-court legal services).

References

[Bh83] Bhatia, Vijay K: Simplification v. easification-the case of legal texts. Applied linguistics, 4:42, 1983.

[BM14] Binns, Reuben; Matthews, David: Community structure for efficient information flow in'ToS; DR', a social machine for parsing legalese. In: Proceedings of the 23rd International Conference on World Wide Web. ACM, pp. 881–884, 2014.

[BMWT14] Bakos, Yannis; Marotta-Wurgler, Florencia; Trossen, David R: Does anyone read the fine print? Consumer attention to standard-form contracts. The Journal of Legal Studies, 43(1):1–35, 2014.

[Bo11] Bollmann, Marcel: Adapting simplenlg to german. In: Proceedings of the 13th European Workshop on Natural Language Generation. pp. 133–138, 2011.

[Br17] Braun, Daniel; Scepankova, Elena; Holl, Patrick; Matthes, Florian: SaToS: Assessing and Summarising Terms of Services from German Webshops. In: Proceedings of the 10th International Conference on Natural Language Generation. Association for Computational Linguistics, Santiago de Compostela, Spain, pp. 223–227, 2017.

[Br18] Braun, Daniel; Scepankova, Elena; Holl, Patrick; Matthes, Florian: Customer-centered LegalTech: Automated Analysis of Standard Form Contracts. In: Tagungsband Internationales Rechtsinformatik Symposium (IRIS) 2018. Editions Weblaw, pp. 627–634, 2018.

[Co15] Collantes, Miguel; Hipe, Maureen; Sorilla, Juan Lorenzo; Tolentino, Laurenz; Samson, Briane: Simpatico: A Text Simplification System for Senate and House Bills. In: Proceedings of the 11th National Natural Language Processing Research Symposium. pp. 26–32, 2015.

[Fr10] Francesconi, Enrico; Montemagni, Simonetta; Peters, Wim; Tiscornia, Daniela: Semantic processing of legal texts: Where the language of law meets the law of language, volume 6036. Springer, 2010.

[GR09] Gatt, Albert; Reiter, Ehud: SimpleNLG: A realisation engine for practical applications. In: Proceedings of the 12th European Workshop on Natural Language Generation (ENLG 2009). pp. 90–93, 2009.

[Gr15] Grabmair, Matthias; Ashley, Kevin D; Chen, Ran; Sureshkumar, Preethi; Wang, Chen; Nyberg, Eric; Walker, Vern R: Introducing LUIMA: an experiment in legal conceptual retrieval of vaccine injury decisions using a UIMA type system and tools. In: Proceedings of the 15th International Conference on Artificial Intelligence and Law. ACM, pp. 69–78, 2015.

[JU17] JURIS Rechtsinformationsdatenbank. `https://www.juris.de/r3/search`, 2017. Accessed: 2017-10-29.

[Li17] Lippi, Marco; Palka, Przemyslaw; Contissa, Giuseppe; Lagioia, Francesca; Micklitz, Hans-Wolfgang; Panagis, Yannis; Sartor, Giovanni; Torroni, Paolo: Automated Detection of Unfair Clauses in Online Consumer Contracts. In: JURIX. pp. 145–154, 2017.

[Ma08] Malaga, Ross A: Worst practices in search engine optimization. Communications of the ACM, 51(12):147–150, 2008.

[Mc05] McCallum, Andrew: Information extraction: Distilling structured data from unstructured text. Queue, 3(9):48–57, 2005.

[MH18] Möllers, Frederik; Hessel, Stefan: Das Sicherheitsgutachten zum besonderen elektronischen Anwaltspostfach (beA). Computer und Recht, 34(7):413–417, 2018.

[Oe06] Oehler, Andreas: Zur ganzheitlichen Konzeption des Verbraucherschutzes - eine ökonomische Perspektive. Verbraucher und Recht, 21(8):294–300, 2006.

[Re08] Reichardt, Vanessa: Der Verbraucher und seine variable Rolle im Wirtschaftsverkehr. Duncker & Humblot, 2008.

[SSW17] Shulayeva, Olga; Siddharthan, Advaith; Wyner, Adam: Recognizing cited facts and principles in legal judgements. Artificial Intelligence and Law, 25(1):107–126, 2017.

[Ta11] Tamm, Marina: Verbraucherschutzrecht: Europäisierung und Materialisierung des deutschen Zivilrechts und die Herausbildung eines Verbraucherschutzprinzips. Mohr Siebeck, 2011.

[Wa17] Waltl, B.; Landthaler, J.; Scepankova, E.; Matthes, F.; Geiger, T.; Stocker, C.; Schneider, C.: Automated extraction of semantic information from german legal documents. In: IRIS: Internationales Rechtsinformatik Symposium. 2017.

[WP06] Walter, Stephan; Pinkal, Manfred: Automatic extraction of definitions from German court decisions. In: Proceedings of the workshop on information extraction beyond the document. Association for Computational Linguistics, pp. 20–28, 2006.

[Ze14] Zerres, Thomas: Principles of the German Law on Standard Terms of Contract. Jurawelt, 2014.

Smart Contracts und die DSGVO

**Welche Grenzen setzt die DSGVO der Verwendung von Smart Contracts?
Eine Betrachtung von Smart Contracts auf der Ethereum-Blockchain**

Jörn Erbguth[1]

Abstract: Über Smart Contracts auf der Ethereum-Blockchain wurden bereits Milliarden-Beträge transferiert. Allerdings wurde bislang wenig betrachtet, ob diese Abwicklung DSGVO-konform war. Dieser Beitrag erörtert, wer Verantwortliche für die Ausführung eines Smart Contracts sind. Dabei fällt auf, dass man hier je nach Gestaltung des Smart Contracts und der Situation der Vertragsparteien zu sehr unterschiedlichen Ergebnissen kommen kann. Darauf aufbauend wird die Verarbeitung personenbezogener Daten und die automatisierte Entscheidung durch einen Smart Contract betrachtet. Auch wenn das Ergebnis abhängig vom Einzelfall ist, so sind der Abschluss und die Abwicklung von Verträgen über Smart Contracts auf der Ethereum-Blockchain prinzipiell DSGVO-konform möglich.

Keywords: Smart Contracts, Blockchain, Ethereum, ADM, automatisierte Entscheidung, DSGVO, personenbezogene Daten, Verantwortlicher

1 Einleitung

Smart Contracts auf öffentlichen Blockchains wie z.B. Ethereum erlauben eine Art Treuhandfunktion. Man kann Verträge eingehen und der Smart Contract wacht darüber, dass z.B. der monetäre Transfer erst endgültig wird, wenn die Leistung auch erbracht ist. Umgekehrt können ggf. Leistungen blockiert werden, wenn die Bezahlung noch nicht erfolgt ist. Dabei bieten Smart Contracts auf der Blockchain die Sicherheit, dass die Bedingungen der Automatisierung nicht einseitig geändert werden können, wie das beispielsweise beim Digital Rights Management (DRM) der Fall ist. Im Folgenden wird betrachtet, ob die DSGVO einen Vertragsabschluss und eine Vertragsdurchführung via Smart Contract auf einer öffentlichen Blockchain wie z.B. Ethereum zulässt. Zur Frage der Vereinbarkeit von Blockchain und DSGVO kommt zudem die Problematik der automatisierten Entscheidung, welche durch Art. 22 DSGVO reguliert ist.[2]

[1] Universität Genf, Institute of Information Service Science, CUI Battelle bat A Route de Drize 7, CH-1227 Carouge Ort, joern@erbguth.net

[2] Bedanken möchte ich mich für wertvolle Anregungen von Mitgliedern der Arbeitsgruppe DIN SPEC 4997 Privacy by Blockchain Design, wobei ich besonders Katrin Kirchert und Michael Kolain hervorheben möchte.

doi:10.18420/inf2019_59

2 Rollen und Begriffe

Vorab sollen einige zentrale Begriffe und Rollen definiert werden. Bei Smart Contracts lassen sich folgende Rollen identifizieren. Dabei können mehrere Rollen in einer natürlichen oder juristischen Person zusammenfallen oder auch noch weiter differenziert werden.

Smart Contracts im Kontext von Blockchains werden hier nicht im weiten Sinne von Szabo[3] als automatisch ausgeführte Verträge verstanden. Vielmehr wird der Begriff auf solche Smart Contracts beschränkt, die Programme auf einer programmierbaren Blockchain sind und die Transaktionen ausführen.[4] Programme auf einer Blockchain, die keine Transaktionen ausführen, sind jedoch nicht mit umfasst.[5]

Die *Entwickler*in*[6] entwickelt den Code des Smart Contracts. Dies umfasst ggf. auch die Spezifikation, die Codierung und das Testen, aber nicht das Deployen auf einer produktiven Blockchain.

Die *Deployer*in* stellt den Smart Contract auf eine produktive Blockchain und macht ihn dadurch einsatzbereit. Danach ist ein Smart Contract an sich unveränderbar. Smart Contracts können jedoch so gebaut werden, dass Updates eingespielt oder sie dauerhaft deaktiviert werden können.

Die *Auftraggeber*in* beauftragt die Entwicklung und Wartung des Smart Contracts. Sie hat dazu Verträge mit der Entwickler*in und der Deployer*in.

Vertragsparteien können über Smart Contracts Verträge schließen oder/und ausführen.[7] Dabei können wie bei anderen juristischen Verträgen die Bedingungen ausgewogen oder auch recht einseitig definiert sein. Vertragsparteien können jedoch nicht den Code des Vertrages ändern. Könnte eine Vertragspartei dies, so wäre sie gleichzeitig Deployer*in.

Orakel sind Dritte, die Informationen an einen Smart Contract liefern und die dieser zur Entscheidung über Transaktionen verwendet. Orakel sind in der Regel keine Vertragsparteien.

Eine *Knotenbetreiber*in* betreibt einen Knoten der Blockchain. Sie führt dabei alle Smart-Contract-Transaktionen aus, speichert den Inhalt der Blockchain und gibt diesen weiter. Sie nimmt jedoch keinen Einfluss auf die Verarbeitung. Würde sie Einfluss auf die Verarbeitung nehmen, so würde sie aus der Blockchain ausgeschlossen werden. Etwas anderes kann im Fall einer *Hard Fork* gelten, bei dem die kollektive Abweichung zur Spaltung einer Blockchain in zwei unabhängige Blockchains führt.

[3] [Sz97] Nick Szabo, first monday, Vol. 2, Nr. 9, 1.9.1997.

[4] [Bu14] Vitalik Buterin et al., Ethereum White Paper.

[5] Siehe dazu etwa [Er18] Erbguth, 33.

[6] Es werden die mit * gegenderten Formen verwendet. Sind die weibliche und männliche Form identisch, so wird der weibliche Artikel ohne Markierung durch ein * verwendet. Gemeint sind immer alle Geschlechter.

[7] Zur Frage, in welchen speziellen Fällen sowohl Vertragsschluss als auch Festlegung des Vertragsinhalts über den Smart Contract selbst getätigt werden können, siehe beispielsweise [Dj16] Djazayeri jurisPR-BKR Anm. 1 E II; [Ka16] Kaulartz, 201 oder [Er19] Erbguth, 26.

Ein *Miner* erstellt neue Blöcke einer Blockchain. Dabei verwendet er beim Proof of Work Rechenleistung im Wettstreit mit anderen Minern. Bei anderen, weniger kompetetiven Konsensverfahren wird die Rolle ggf. auch *Blockproducer* genannt. Für die Zwecke dieses Aufsatzes wird allgemein der Begriff Miner verwendet. Charakteristisch für Miner ist, dass diese zwar einen kleinen Einfluss auf die Reihenfolge der Transaktionen einer Blockchain haben, ansonsten aber isoliert keinen Einfluss auf Blockchains ausüben können.

Neuere Blockchains wie z.B. EOS haben einen eingebauten Mechanismus zur *Dispute Resolution* und *Governance*.[8] Die Governance steuert dabei allgemein die Weiterentwicklung des Systems, die Änderung von Regeln und bei der On-Chain-Governance auch deren direkte Umsetzung. Eine solche Governance kann auch in Smart Contracts eingebaut werden. Zusätzlich kann auch eine Dispute Resolution ähnlich eines Schiedsgerichtsverfahrens in eine Blockchain oder einen Smart Contract eingebaut werden, um Einzelfälle zu entscheiden. Governance und Dispute Resolution können auch auf eine verbundene Blockchain ausgelagert werden.[9] Die Dispute Resolution wird dabei nicht von sich aus tätig, sondern muss von einer Vertragspartei angerufen werden. Die Umsetzung der Entscheidungen kann dabei im Smart Contract programmiert sein.

3 Abbildung auf die Rollen der DSGVO

Die DSGVO kennt die Rollen der Verantwortlichen (Art. 4 Nr. 7 DSGVO), der gemeinsam Verantwortlichen (Art. 26 Abs. 1 DSGVO), der Auftragsverarbeiter*in (Art. 4 Nr. 8 DSGVO) sowie der Betroffenen. Die Abbildung dieser Rollen auf die im Kontext eines Smart Contracts vorhandenen Rollen ist abhängig von der konkreten tatsächlichen, technischen und rechtlichen Gestaltung. Da bei der Frage der Anwendbarkeit der DSGVO auch auf diese Rollen abgestellt wird, müssen diese Rollen vorab geklärt werden.

3.1 Verantwortliche und Auftragsverarbeiter*innen

Verantwortliche bestimmen die Mittel und Zwecke der Datenverarbeitung (Art. 4 Nr. 7 DSGVO). Im englischen werden sie als *controller* bezeichnet, was deutlich macht, dass hier neben der rechtlichen Verantwortlichkeit der faktische Einfluss wichtig ist. Wer weder rechtlich noch tatsächlich Einfluss auf die Entscheidung über die Verarbeitung hat, kann nicht als für die Verarbeitung Verantwortliche angesehen werden.[10] Wer im Auftrag der Verantwortlichen eine Verarbeitung durchführt, ist Auftragsverarbeiter*in (Art. 4 Nr. 8 DSGVO). Die französische Datenschutzaufsichtsbehörde CNIL hat in einer Stellungnahme[11] erwogen, Entwickler*innen als Auftragsverarbeiter*innen oder Verantwortliche anzusehen.

[8] ECAF, The EOS-Core Arbitration Forum, https://www.eoscorearbitration.io/
[9] [KW17] Kolain/Wirth.
[10] [Ar10] Artikel-29-Datenschutzgruppe WP169, 15.
[11] [Cn18] CNIL, 4.

Dabei hat die CNIL jedoch eingeschränkt, dass dies nur gelte, wenn sie Einfluss auf die Verarbeitung nehmen. Wenn sie keine weiteren Rollen, wie etwa die der Deployer*in übernehmen, ist dies jedoch nicht der Fall.

Deployer*innen stellen den Smart Contract zur Nutzung bereit. Dabei muss unterschieden werden, ob sie die Kontrolle behalten oder aufgeben. Es gibt dabei insgesamt fünf Fälle:

1. Die Deployer*in gibt die Kontrolle unmittelbar ab. Dies ist der klassische Fall eines Smart Contracts auf einer Blockchain. Einmal geschrieben ist er final und kann nicht aktualisiert oder deaktiviert werden.
2. Die Deployer*in behält die Kontrolle und kann über ihre privaten Schlüssel Updates einspielen oder den Smart Contract deaktivieren.
3. Die Deployer*in behielt beim ursprünglichen Deployment zwar die Kontrolle, deaktiviert die Kontrolle jedoch endgültig, bevor Vertragsparteien den Smart Contract verwenden.
4. Die Deployer*in deaktiviert die Kontrolle erst, nachdem Vertragsparteien den Smart Contract zu verwenden begonnen haben.
5. Die Deployer*in deaktiviert die Kontrolle in Absprache mit den Vertragsparteien, nachdem diese den Smart Contract zu verwenden begonnen haben oder gibt die Kontrolle in Absprache mit den Vertragsparteien an eine Dritte ab.

Im ersten Fall hat die Deployer*in keine Kontrolle darüber, welche Vertragsparteien den Smart Contract wofür verwenden. Sie kann die Verarbeitung auch nicht mehr beeinflussen. Sie hat lediglich den Code bereitgestellt, den andere dann verwenden. Wer den Code verwenden will, muss für die Ausführung bezahlen. Die Bezahlung für die Ausführung geht dabei an den Miner des entsprechenden Blocks, in dem die Ausführung festgehalten wird. Die CNIL sieht Smart-Contract-Entwickler*innen als Auftragsverarbeiter*innen, wenn sie nicht nur Code bereitstellen, sondern die Ausführung des Codes beeinflussen.[12] Weder die konkrete Ausführung des Codes noch ob der Code überhaupt (und wenn dann vom wem) ausgeführt wird, kann von der Deployer*in beeinflusst werden. Wer den Code ausführen möchte, kann ihn sich zudem vorab ansehen.[13]

Der dritte Fall ist dem ersten Fall gleichzustellen, da die Deployer*in die Kontrolle bereits aufgegeben hat, bevor der Smart Contract Verwendung findet. Behält sie dagegen wie im zweiten Fall die Kontrolle und kann die Bearbeitung weiterhin beeinflussen, so kann sie die Verarbeitung steuern und ist als Verantwortliche – oder falls sie dies im Auftrag macht, als Auftragsverarbeiter*in – einzustufen. Fraglich ist besonders der vierte Fall: Kann eine Deployer*in als Verantwortliche eingestuft werden, wenn sie keine Kontrolle mehr hat?

[12] [Cn18] CNIL, 2.
[13] [Er19] Erbguth, 29

Sie war Verantwortliche, als die Vertragsparteien mit der Verwendung des Smart Contracts begonnen haben. Entledigt sie sich dieser Kontrolle, können die Vertragsparteien deshalb nicht schutzlos gestellt werden und sie bleibt in der Verantwortung. Etwas anderes gilt im fünften Fall, falls eine Aufgabe oder Abgabe der Kontrolle und Verantwortung in Absprache mit den Vertragsparteien erfolgt. Im Ergebnis kann die Deployer*in Verantwortliche sein, wenn sie die Kontrolle behält oder sie beim Beginn der Verwendung hatte und sie behalten hätte sollen.

Die Akteur*innen der Blockchain, Knotenbetreiber*innen und Miner, könnten ebenfalls Verantwortliche der Ausführung des Smart Contracts sein. Dabei sind die gleichen Kriterien anzuwenden, die für die Blockchain gelten. Knotenbetreiber*innen und Miner haben demnach bei öffentlichen Blockchains in der Regel keine Kontrolle. Sie können jedoch als Auftragsverarbeiter*innen klassifiziert werden.[14]

Die Vertragsparteien entscheiden darüber, den Smart Contract verwenden zu wollen. Ist der Smart Contract unveränderbar, so sind die Vertragsparteien die Einzigen, die auf diese Entscheidung Einfluss haben. Daher entscheiden sie auch über die Zwecke und Mittel der Datenverarbeitung.[15] Gibt es bei den Vertragsparteien ein Machtgefälle, so kann sich die Verantwortlichkeit ggf. auf eine der beiden Vertragsparteien beschränken. Dies ist etwa der Fall, wenn der Smart Contract der einen Vertragspartei deutlich mehr Aktionsmöglichkeit gibt. Das gleiche kann gelten, wenn eine Vertragspartei direkt oder indirekt Einfluss auf die Entwicklung des Smart Contract genommen hat und damit die Bedingungen „diktiert“.

Bei Smart Contracts werden häufig unparteiische Dritte als Orakel eingebunden, die Informationen beisteuern, welche die Ausführung des Smart Contract maßgeblich beeinflussen. So muss etwa ein Smart Contract zur Versicherung von Flugverspätungen auf die geplanten und tatsächlichen Ankunftszeiten der Flüge zugreifen. Diese Information muss von einer vertrauenswürdigen Dritten bereitgestellt werden. Auf der einen Seite stellt die Dritte hier "nur" Informationen bereit. Auf der anderen Seite ist die Dritte hier möglicherweise die Einzige, die hier noch einen Einfluss auf die Auszahlung der Flugverspätung nehmen kann. Es ist fraglich, ob dies für die Annahme einer Verantwortlichkeit ausreicht. Solange sie diesen Einfluss jedoch in der von den Vertragsparteien vorgegebenen Art und Weise ausübt, macht sie dies allenfalls als Auftragsverarbeiter*in und ist nicht selbst Verantwortliche.

Smart Contracts können auch Funktionen zur Governance und zur Dispute Resolution eingebaut haben. Diese können die Ausführung des Smart Contracts ggf. blockieren oder auch verändern. Dabei muss unterschieden werden zwischen einer Dispute Resolution, die erst auf Initiative einer oder mehrerer Vertragsparteien aktiviert wird und den Eingriffsmöglichkeiten der Governance, die kein Zutun der Vertragsparteien erfordern. Im ersten Fall entsteht die Kontrolle erst dadurch, dass eine oder mehrere Vertragsparteien die Dispute

[14] So die CNIL in [Cn18], 2; davor bereits [MW17] Martini/Weinzierl, 1250; [EF17] Erbguth/Fasching, 563; ebenso Janicki/Saive mit weiteren Nachweisen [JS19], 253; Schrey/Thalhofer sehen dagegen alle Teilnehmer*innen einer Blockchain als Verantwortliche [ST17], 1433

[15] In Fortführung der Argumentation der CNIL in [CN18], 2.

Resolution anrufen. Im letzteren Fall besteht jedoch eine ständige Kontrollmöglichkeit und damit möglicherweise eine Verantwortlichkeit. Charakteristisch ist dabei auch, dass die Governance und Dispute Resolution nicht weisungsgebunden sind. Allerdings muss die Governance den vorgegebenen Governanceregeln folgen. Das alleine macht sie jedoch noch nicht zur Auftragsverarbeiter*in. Sofern die Governance nicht durch eine übergeordnete Auftraggeber*in kontrolliert wird, dürfte die Governance damit die letztendliche Instanz sein, die über die Mittel und Zwecke der Datenverarbeitung entscheidet. Fraglich ist jedoch, ob es angemessen ist, private Schieds- und Überwachungsinstanzen als Verantwortliche anzusehen. Gerade durch den Einbau solcher Instanzen soll die Compliance mit den vertraglichen und gesetzlichen Regeln sichergestellt werden. Aber genau der Einbau solcher Kontrollmöglichkeiten erhöht die Verantwortlichkeit. Das entspricht jedoch dem der DSGVO innewohnenden Prinzips, dass die Verantwortlichkeit dort angesiedelt wird, wo auch tatsächlich entschieden wird.

3.2 Gemeinsame Verantwortlichkeit

Je nach Konstellation können wie zuvor dargestellt neben den Vertragsparteien verschiedene weitere Akteur*innen Verantwortliche sein. Fraglich ist dabei, ob alle zusammen eine gemeinsame Verantwortlichkeit trifft oder ob bestimmte Verantwortlichkeiten auf Grund anderer, dominierender Verantwortlichkeiten zurücktreten. In C-210/16 hat der EuGH entschieden, dass gemeinsame Verantwortlichkeiten auch bestehen können, wenn die Verantwortlichkeit nicht gleich verteilt ist.[16] Ist der Beitrag einer Akteur*in jedoch verschwindend gering im Vergleich zur dominierenden Kontrolle anderer, so erscheint die Annahme gemeinsamer Verantwortlichkeit nicht angemessen.

Hier muss abgewogen werden:

- Hat eine Vertragspartei die Entwicklung und das Deployment beauftragt und kontrolliert damit die Ausführung des Smart Contract, so tritt demgegenüber die Kontrollmöglichkeit der anderen Vertragsparteien in den Hintergrund.
- Gibt es keine Kontrolle durch das Deployment, so haben die Vertragsparteien die alleinige Kontrolle. Ist hier kein starkes Machtgefälle zwischen den Vertragsparteien erkennbar, so spricht viel für eine gemeinsame Verantwortlichkeit.
- Hat nach Start des Smart Contracts durch die Vertragspartner*innen nur noch die Governance Kontrolle über die Ausführung des Smart Contracts, so ist die Governance als Verantwortliche zu qualifizieren.
- Ist eine Verantwortliche gleichzeitig Betroffene, so muss sie die Begrenzungen der DSGVO bzgl. der Verarbeitung ihrer eigenen Daten und bzgl. der sie betreffenden automatisierten Entscheidungen nicht beachten.

[16] EuGH Urteil vom 5. Juni 2018, C-210/16, Rn 43

3.3 Betroffene

Betroffene sind die Personen, auf die sich die personenbezogenen Daten beziehen. Bei automatisierten Entscheidungen (Artikel 22 DSGVO) sind Betroffene auch diejenigen, die einer ausschließlich auf einer automatisierten Verarbeitung beruhenden Entscheidung unterworfen sind.

4 Anwendbarkeit der DSGVO

4.1 Sachlicher Anwendungsbereich

Die Definition von personenbezogenen Daten in Art. 4 Nr. 1 der DSGVO ist sehr weit. Sind Daten für diejenigen, die Zugriff auf die Daten haben können, mit einer natürlichen Person verknüpfbar, so dass Information zu diesen Personen abgeleitet werden können, dann handelt es sich bereits um personenbezogene Daten. Dabei müssen sämtliche Techniken und auch Daten einbezogen werden, die dazu herangezogen werden können. Der Erwägungsgrund 26 schränkt dies insofern ein, als nur rein theoretische Möglichkeiten ausgeschlossen werden. In C-210/16 legt der EuGH die Schwelle für eine Identifizierbarkeit recht hoch, in dem er eine Identifizierungsmöglichkeit bereits dann außer Acht lassen möchte, bei der das Risiko einer Identifizierung "de facto"vernachlässigbar wäre, da die Identifizierung verboten ist oder einen unverhältnismäßigen Aufwand erfordert.[17] Der Erwägungsgrund stellt dabei nicht nur auf die Verantwortliche ab, sondern bezieht auch *andere Personen* mit ein, die über Heranziehung weiterer Informationen oder sonstiger Mittel, natürliche Personen mit den Daten zu identifizieren. Das ist aber nur möglich, wenn diese andere Personen auch Zugriff auf die in Frage stehenden Daten erlangen könnten.

Privacy Enhancing Technology, also Techniken wie Verschlüsselung, kryptographische Hashwerte oder Zero Knowledge Proofs sind Werkzeuge für Datenschutz durch Technik, deren Verwendung in Art. 25 DSGVO eingefordert wird. Wegen eingeschränkter organisatorischer Schutzmöglichkeiten wird im Kontext öffentlichen Blockchains Datenschutz vor allem durch Technik sichergestellt. In einem speziellen Anwendungsfall der versuchten Anonymisierung von Datensätzen haben die Aufsichtsbehörden 2014 beschrieben, dass eine Anonymisierung durch Verschlüsseln, Hashing oder Löschen der direkt personenbezogenen Merkmale in der Regel nicht erfolgreich möglich ist.[18] Damit sind jedoch Techniken wie Hashing nicht generell ungeeignet, um auf Blockchains eingesetzt zu werden. Vielmehr muss im Einzelfall geprüft werden, wie die Technik eingesetzt wird und ob ein Personenbezug nach den in C-210/16 und im Erwägungsgrund 26 beschriebenen Kriterien noch hergestellt werden kann.[19] Dafür spricht auch eine Entscheidung der österreichischen Datenschutzbehörde: In diesem Fall, der weder Blockchain noch Smart Contracts betraf, stellt die Behörde

[17] EuGH, Urteil vom 19. Oktober 2016, C-582/14, Rn 46.
[18] [Ar14] Artikel-29-Datenschutzgruppe WP216, 29.
[19] so auch das EU Blockchain Observatory [Bl18], 22; a.A. Finck [Fi18], 22

fest, dass ein effektiv nicht mit einer Person in Verbindung zu bringender Datensatz als anonymisiert gilt und damit einem Löschen gleichkommt.[20] Erwägungsgrund 26 sieht allerdings auch vor, dass über die zum Zeitpunkt der Verarbeitung verfügbare Technologie hinaus auch technologische Entwicklungen zu berücksichtigen sind. Darunter könnte der Zuwachs an Rechenleistung nach dem Mooreschen "Gesetz"[21] oder Quantencomputer[22] fallen.

Sofern Bitcoin- oder Ethereum-Transaktionen mit einer öffentlichen Adresse einer Privatperson verknüpft sind, dürfte hier analog zu den IP-Adressen ein personenbezogenes Datum vorliegen.[23] Für Bitcoin gibt es beispielsweise kommerzielle Anbieter, die die Personen identifizieren, die hinter einer Adresse stehen.[24]

4.2 Haushaltsausnahme

Art. 2 Abs. 2 lit. c DSGVO stellt Datenverarbeitungen von der Anwendung der DSGVO frei, soweit die Datenverarbeitung ausschließlich zur Ausübung persönlicher oder familiärer Tätigkeiten dient. In C-101/01[25] hat der EuGH in Bezug auf für die Öffentlichkeit zugängliche Information festgestellt, dass dies den Bereich der Haushaltsausnahme überschreite. Hieran hat der EuGH kürzlich in C345/17[26] ausdrücklich festgehalten. Mit dem Schreiben eines Eintrags auf eine öffentliche Blockchain ist der Eintrag nicht nur öffentlich zugreifbar, sondern wird auch zehntausendfach kopiert.[27] Die CNIL hat dagegen angenommen, dass eine privat motivierte Transaktion auf einer öffentlichen Blockchain unter die Haushaltsausnahme fällt.[28] Allerdings sind eine Blockchain-Transaktion und eine Äußerung im Web oder Social Media nicht unbedingt vergleichbar. Die Transaktion ist nur mit größeren Aufwand zuordenbar. Eine Öffentlichkeit wird damit meistens effektiv nicht hergestellt. Zudem würde eine enge Interpretation der Haushaltsausnahme dazu führen, dass Privatpersonen mit ihren Transaktionen auf Blockchains selbst zu Verantwortlichen werden. Sie müssten sich gegenüber den Betroffenen identifizieren. Die DSGVO würde damit z.B. "anonyme" Bitcoin-Zahlungen verbieten. Ohne hier näher auf die umfangreiche Diskussion zu den Schutzzwecken der DSGVO[29] einzugehen, wäre das wohl eine Konsequenz, die möglicher-

[20] DSB-D123.270/0009-DSB/2018.

[21] Das Mooresche "Gesetz"beschreibt die Beobachtung, dass neue Rechner alle 2 Jahre etwa doppelt so leistungsfähig sind. Dies bedeutet, dass Rechner in 40 Jahren etwa eine Million mal schneller sein werden.

[22] Quantencomputer rechnen nicht schneller, sondern anders und können dadurch Dinge berechnen, die konventionell nur durch langwieriges Ausprobieren ermittelbar wären. Eine Abschätzung, welche kryptographische Technik dadurch unsicher wird und welche sicher bleibt, gibt es bei [Ch16] Chen et al., 2.

[23] [EF17] Erbguth/Fasching, 561, a.A. [ST17] Schrey/Thalhofer, 1433.

[24] So etwa Chainalysis.com 01.05.2019.

[25] EuGH, Urteil vom 06.11.2003, C-101/01, Rn 47.

[26] EuGH, Urteil vom 14.02.2019, C-345/17, Rn 41.

[27] Die Anzahl der Knoten ist etwa über `https://www.ethernodes.org` sichtbar.

[28] [Cn18] CNIL, 2; a.A. Janicki/Saive [JS19], 252.

[29] So etwa die Diskussion von Winfried Veil und Kirsten Bock auf dem CRonline-Blog vom 6.2.2019, 18.3.2019, 22.3.2019 und 29.3.2019 `https://www.cr-online.de/blog/01.05.2019`.

weise nicht dem Normzweck der DSGVO entspricht. Dies stützt die Position der CNIL und spricht für eine etwas weitere Interpretation der Haushaltsausnahme.

4.3 Räumlicher Anwendungsbereich

Der räumliche Geltungsbereich wird in Art. 3 DSGVO geregelt. Dabei wird nicht auf den Ort der Datenverarbeitung abgestellt, vielmehr genügt, wenn eines der folgenden Merkmale einschlägig ist:

- Die Verarbeitung erfolgt im Rahmen der Niederlassung einer Verantwortlichen oder einer Auftragsverarbeiter*in in der Europäischen Union.
- Die Datenverarbeitung steht im Zusammenhang mit dem Angebot von Waren oder Dienstleistungen an Personen in der Europäischen Union.
- Die Datenverarbeitung steht im Zusammenhang mit der Beobachtung des Verhaltens von Personen in der Europäischen Union.

Bei einer öffentlichen Blockchain mit an die 10.000 Knoten sitzen sicher auch Knotenbetreiber in der EU, so dass daher die DSGVO räumlich Anwendung findet. Dies gilt selbst dann, wenn die anderen Akteure, wie etwa die Vertragsparteien, nicht in der EU ansässig sind.

5 Rechtfertigung zur Verarbeitung personenbezogener Daten

Bei Smart Contracts muss zwischen rein lesenden Aufrufen und Transaktionen unterschieden werden. Nur bei Letzteren wird die Verarbeitung durch jeden Knoten wiederholt und dabei Input und Output sowie Zustandsänderung auf der Blockchain abgespeichert. Bei Ethereum stehen diese Inhalte danach unveränderlich auf der öffentlichen Blockchain. Es stellen sich daher folgende Fragen:

1. Sind Ein- und Ausgabedaten sowie abgelegte Zustandsdaten personenbezogene Daten? Dies muss auf Grund der Daten beurteilt werden. Da die Definition von personenbezogenen Daten recht weit gezogen ist, dürfte dies häufig der Fall sein.
2. Wer sind Betroffene und wer Verantwortliche dieser Daten. Unproblematisch ist es dabei, wenn Verantwortliche mit Hilfe eines Smart Contracts eigene personenbezogene Daten verarbeiten.
   Gibt es jedoch keine Personenidentität zwischen Verantwortlichen und Betroffenen und ist zudem die Haushaltsausnahme nicht einschlägig, so ist eine Rechtfertigung für diese Verarbeitung erforderlich.

5.1 Einwilligung (Art. 6 Abs. 1 S. 1 lit. a DSGVO)

Nach Art. 6 Abs. 1 lit. a DSGVO könnten die Betroffenen in die Verarbeitung der personenbezogenen Daten einwilligen. Problematisch daran ist, dass die Einwilligung jederzeit widerrufbar ist (Art. 7 Abs. 2 S. 1 DSGVO). Der Widerruf wirkt jedoch nur für künftige Verarbeitungen (Art. 7 Abs. 3 S. 2 DSGVO). Die Rechtmäßigkeit der vor dem Widerruf erfolgten Verarbeitung ist daher davon nicht betroffen. Die Transaktion muss dementsprechend nicht rückgängig gemacht werden. Allerdings bleiben die Daten auch danach noch gespeichert, was nach Art. 4 Nr. 2 DSGVO ebenfalls eine Verarbeitung ist. Die Speicherung ist logische Folge der Verarbeitung des Smart Contracts. Selbst wer Kontrolle über den Smart Contract, dessen Deaktivierung oder ein Update hat, kann die Daten nicht entfernen. Daher muss die Speicherung der für die Vornahme der Transaktion Verantwortlichen zugerechnet werden. Fraglich ist, ob dies bedeutet, dass die dauerhafte Speicherung auch als direkter statischer Erfolg der Transaktion des Smart Contracts gilt und daher beim Widerruf der Einwilligung nicht rückgängig gemacht werden muss. Für eine solche Auslegung spricht, dass nach Einwilligung genau das getan wurde, worin eingewilligt wurde und der Widerruf eben das nicht rückabwickeln soll. Damit wäre jedoch das Recht auf Widerruf und im Endeffekt auch das Recht auf Vergessenwerden (Art. 17 DSGVO) deutlich eingeschränkt. Selbst wenn man dieser Überlegung folgte, so müsste das zumindest auf Fälle begrenzt sein, in denen sich die dauerhafte Speicherung zwingend ergibt, in diesem Kontext auch vom Betroffenen gewollt war und in denen vor der Einwilligung sehr deutlich auf diesen Umstand hingewiesen wurde. Keineswegs sollten Unternehmen sich dadurch der Löschpflicht entziehen können, indem ihre Verfahren keine Löschmöglichkeit vorsehen. Im Ergebnis bleibt die Einwilligung daher ein eher ungeeignetes Instrument der Rechtfertigung.

5.2 Erfüllung eines Vertrages (Art. 6 Abs. 1 S. 1 lit. b DSGVO)

Erfolgt die Verarbeitung zur Erfüllung eines Vertrags, dessen Vertragspartei die betroffene Person ist, könnte die Verarbeitung zulässig sein. Dies gilt auch, wenn die Verarbeitung für vorvertragliche Maßnahmen erforderlich ist. Die Erforderlichkeit wurde vom EDSA näher ausgeführt.[30] Sofern der Vertrag originär mit dem Smart Contract begründet wird oder zumindest ein anderweitig geschlossener Vertrag eine Ausführung als Smart Contract auf einer Blockchain erfordert, kann das hier einschlägig sein. Allerdings können nicht beliebige Sachverhalte durch Implementierung als Smart Contract über Art. 6 Abs. 1 S. 1 lit. b DSGVO datenschutzrechtlich gerechtfertigt werden. Vielmehr muss der Sachverhalt an sich eine Implementierung als Smart Contract mit dauerhafter Speicherung rechtfertigen.

[30] [Ed19] EDPB: Guidelines 2/2019, Rn 13.

5.3 Rechtliche Verpflichtung (Art. 6 Abs. 1 S. 1 lit. c DSGVO)

Sofern die Verarbeitung zur Erfüllung einer rechtlichen Verpflichtung erforderlich ist, der die Verantwortliche unterliegt, ist die Verarbeitung ebenfalls gerechtfertigt. Sofern es Verpflichtungen gibt, Transaktionen dauerhaft einsehbar und überprüfbar zu machen, können Smart Contracts ein gutes Hilfsmittel sein. Ist dagegen eine solche Transparenzpflicht zeitlich begrenzt, so müsste eine spezielle Blockchain gebaut werden, die Daten ebenfalls nur genau so lange speichert.

5.4 Berechtigtes Interesse (Art. 6 Abs.1 S. 1 lit. f DSGVO)

Im Fall von berechtigtem Interesse, dem kein überwiegendes Interesse der Betroffenen entgegen steht, ist eine Verarbeitung ebenfalls zulässig. Das berechtigte Interesse kann allerdings durch einen Widerspruch der Betroffenen nach Art. 21 Abs. 1 DSGVO auf den Fall zwingender schutzwürdiger Gründe beschränkt werden, die die Interessen der Betroffenen überwiegen. Ein so starkes legitimes Interesse könnte man sich etwa vorstellen, wenn durch Maßnahmen des technischen Datenschutz nur ein sehr geringes Restrisiko für die Rechte der Betroffenen verbleibt.[31] Solch zwingende schutzwürdige Gründe sind ebenfalls denkbar, wenn ein Smart Contract als Vertrag unwirksam ist, zurückabgewickelt werden muss und die Einträge auf der Blockchain zur Richtigstellung der Einträge erforderlich sind. Im Kontext der Unveränderbarkeit werden zwingende schutzwürdige Interessen jedoch nur in seltenen Fällen die Interessen der Betroffenen dauerhaft überwiegen, so dass Art. 6 Abs. 1 S. 1 lit. f DSGVO als generelles Rechtfertigungsinstrument wenig geeignet ist.

6 Automatisierte Entscheidung

Art. 22 DSGVO gibt Betroffenen das Recht, nicht einer Entscheidung unterworfen zu werden, die ausschließlich auf einer automatisierten Verarbeitung personenbezogener Daten beruht.

6.1 Entscheidung

Smart Contracts arbeiten an sich vollständig automatisiert. Doch beruht die Entscheidung über eine Transaktion eines Smart Contracts auf dessen automatisierter Verarbeitung oder aber auf den vorab manuell festgelegten sehr einfachen Regeln? Ist die eigentlich Entscheidung nicht bereits bei der Festlegung dieser einfachen Regeln bzw. bei der Auswahl des transparenten Smart Contracts zur Abwicklung einer Transaktion getroffen worden? Dies würde im Endeffekt bedeuten, dass Art. 22 nur dann Anwendung finden würde, wenn die

[31] [Cn18] CNIL, Premiers éléments d'analyse de la CNIL, 6.

Entscheidungsregeln nicht manuell vorgegeben wurden. Dies ist etwa bei Blackbox-Verfahren der KI der Fall. *Finck* lehnt eine solche restriktive Interpretation mit dem Argument ab, dass diese nicht vom Gesetzgeber gemeint gewesen sein könne. Denn bei dieser Auslegung würde die Ausnahme für *Entscheidungen zur Erfüllung eines Vertrags* in Art. 22 Abs. 2 lit. a DSGVO keinen Sinn mehr machen, da bei manuell geschlossenen Verträgen die Bedingungen einer Transaktion vorab festlegt würden.[32] Gegen eine restriktive Interpretation spricht auch das von der Artikel 29-Gruppe genannte Beispiel der automatisierten Erstellung von Bußgeldbescheiden auf Basis der Messwerte einer Geschwindigkeitsüberwachung. Auch dort sind die Entscheidungskriterien einfach und vorab festgelegt.[33]

Art. 22 DSGVO beschränkt das Verbot auf Entscheidungen, die gegenüber der Betroffenen eine *rechtliche Wirkung* entfalten oder sie *in ähnlicher Weise erheblich beeinträchtigt*. Mit Smart Contracts werden häufig Verträge abgeschlossen, erfüllt und/oder Assets auf der zugrunde liegenden Blockchain verschoben. Da wird eine rechtliche Wirkung meistens gegeben sein.[34] Zudem relativiert die Artikel-29-Gruppe die Begrenzung auf *erhebliche Beeinträchtigungen* so weit, dass selbst die Auswahl von Werbung darunter fallen könnte.[35]

Diese weite Interpretation der *Entscheidung* sollte nicht unkritisch gesehen werden. Denn sie könnte dazu führen, dass nicht nur Transaktionen von Smart Contracts, sondern dass jedwede Ethereum-Blockchain-Transaktion unter Artikel 22 subsumiert werden kann, da dabei Berechtigungen und Kontostände geprüft sowie Kryptocoins bewegt werden.

6.2 Ausnahmen

Art. 22 DSGVO nennt in Absatz 2 mehrere Ausnahmen, die automatisierte Entscheidungen trotzdem erlauben. Hier sind insbesondere Abschluss oder Erfüllung eines Vertrages zwischen Verantwortlichen und Betroffenen (lit. a) sowie die ausdrückliche Einwilligung der Betroffenen (lit. b) zu nennen. Bei Smart Contracts dürfte häufig ein Vertrag abgeschlossen oder erfüllt werden. Da der Umgang mit Smart Contracts auf einer Blockchain meistens bewusst gewählt wird, wird eine ausdrückliche Einwilligung häufig vorliegen oder erteilt werden können.

6.3 Recht auf manuelles Eingreifen, Anhörung und Anfechtung

Art. 22 Abs. 3 gibt in den Fällen der beiden Ausnahmen nach Abs. 2 lit. a und lit. c den Betroffenen das Recht auf ein manuelles Eingreifen, bei dem sie von den Verantwortlichen angehört werden und die Entscheidung anfechten können. Dieses Recht könnte z.B. durch

[32] [Fi19] Finck, 8.
[33] [Ar17] Artikel-29-Gruppe, WP251, 8.
[34] [Fi19] Finck, 10.
[35] [Ar17] Artikel-29-Gruppe, WP251, 23.

eine in den Smart Contract eingebaute Dispute Resolution gewährleistet werden. Hierfür wird beispielsweise eine unabhängige Vertrauensperson kontaktiert, die bei Meinungsverschiedenheiten zwischen den Vertragsparteien eingreifen kann. Ist eine Dispute Resolution weder in den Smart Contract noch der Blockchain an sich eingebaut, ist die Ausführung des Smart Contracts final. Das manuelle Eingreifen kann jedoch auch nachträglich erfolgen.[36] Allerdings muss auch effektiv ein Eingreifen möglich sein, d.h. eine falsche Entscheidung muss korrigiert oder zumindest kompensiert werden können. Etwas unpassend erscheint in Art. 22 Abs. 3 die Anforderung, dass die Verantwortliche die entsprechenden Maßnahmen treffen und eine Person seitens der Verantwortlichen eingreifen muss. Reicht es nicht, sie letztendlich dafür verantwortlich zu machen, dass die Maßnahmen zur Verfügung stehen? Eine neutrale Entscheidungsinstanz sollte hier doch keinen Nachteil darzustellen.

6.4 Ergebnis

Sofern auf Grund der Rollenverteilung die DSGVO für eine Smart-Contract-Transaktion Anwendung findet, muss ein manuelles Eingreifen möglich sein. Dieses Eingreifen muss jedoch die Transaktion auf der Blockchain nicht komplett ungeschehen machen, sondern es reicht, diese außerhalb der Blockchain rückabzuwickeln oder kompensieren zu können.

7 Fazit

Bereits mit der öffentlichen Ethereum-Blockchain sind eine Vielzahl sehr unterschiedlicher Rollenverteilungen möglich, die im Detail zu deutlich unterschiedlichen Ergebnissen führen können. Nicht jede Konstellation ist dabei DSGVO-konform. Im Ergebnis sind z.B. zwei Arten von Smart Contracts DSGVO-konform: Zum einen komplett autonome Smart Contracts z.B. als dezentralisierte Börsen zum Austausch von Tokens. Hier sind die Bedingungen transparent festgelegt und die Vertragsparteien sind die Verantwortlichen i.S.d. DSGVO, in dem sie direkt mit diesem Mechanismus interagieren, sofern nicht die Haushaltsausnahme die Anwendung der DSGVO ausschließt. Zum anderen kontrollierte Smart Contracts bei denen die Verantwortlichen Mechanismen zur Dispute Resolution eingebaut haben oder externe Kompensationsmechanismen für falsche Entscheidungen bereitstellen.

In der Regel unzulässig dürfte es dagegen sein, einen anderweitig geschlossenen Vertrag ohne zwingende Erfordernis einseitig zur Ausführung auf eine Ethereum-Blockchain zu übertragen.

Literatur

[Ar10] Artikel-29-Datenschutzgruppe WP169, 0264/10/DE: Stellungnahme 1/2010 zu den Begriffen "für die Verarbeitung Verantwortlicher" und "Auftragsverarbeiter", 16.02.2010.

[36] [Fi19] Finck, 16.

[Ar14] Artikel-29-Datenschutzgruppe WP216, 0829/14/DE: Stellungnahme 5/2014 zu Anonymisierungstechniken, 10.04.2014.

[Ar17] Artikel 29 Data Protection Working Party: Guidelines on Automated individual decision-making and Profiling for the purposes of Regulation 2016/679, 03.10.2017.

[Bl18] The European Union Blockchain Observatory and Forum, Blockchain and the GDPR, 16.10.2018, https://www.eublockchainforum.eu/sites/default/files/reports/20181016_report_gdpr.pdf[37].

[Bu14] Buterin V.: Etherum White Paper, A next-generation smart contract and decentralized application platform, https://www.weusecoins.com/assets/pdf/library/Ethereum_white_paper-a_next_generation_smart_contract_and_decentralized_application_platform-vitalik-buterin.pdf.

[Ch16] Chen L. et al. Report on Post-Quantum Cryptography, NISTIR 8104, https://nvlpubs.nist.gov/nistpubs/ir/2016/nist.ir.8105.pdf.

[Cn18] CNIL: La blockchain: Premiers éléments d'analyse de la CNIL, 9/2018, https://www.cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf englische Version https://www.cnil.fr/sites/default/files/atoms/files/blockchain.pdf.

[Dj16] Djazayeri, A.: Rechtliche Herausforderungen durch Smart Contracts, jurisPR-BKR 12/2016.

[Ed19] edpb: 'Guidelines 2/2019 on the processing of personal data under Article 6(1)(b) GDPR in the context of the provision of online services to data subjects' 09.04.2019.

[EF17] Erbguth, J.; Fasching J.: Wer ist Verantwortlicher einer Bitcoin-Transaktion? ZD 2017, 560-565.

[Er18] Erbguth, J.: 'Was sind Smart Contracts, wofür werden sie eingesetzt und wann gilt Code is Law?' Telemedicus Sommerkonferenz 2018.

[Er19] Erbguth, J.: 'Transparenz von Smart Contracts' in: Smart Contracts, hrsg. Fries, M.; Paal, B. 2019, Mohr-Siebeck, ISBN 978-3-16-156910-4.

[Fi18] Finck, M.: Blockchains and Data Protection in the European Union, EDPL 2018, 17.

[Fi19] Finck, M.: Smart Contracts as a Form of Solely Automated Processing Under the GDPR, 08.01.2019, Max Planck Institute for Innovation & Competition Research Paper No. 19-01. http://dx.doi.org/10.2139/ssrn.3311370.

[JS19] Janicki, T.; Saive, D.: Privacy by Design in Blockchain-Netzwerken, ZD 2019, 251.

[Ka16] Kaulartz, M.: Herausforderungen bei der Gestaltung von Smart Contracts, InTer 2016, 201.

[KW17] Kolain M.; Wirth C.: Multi-Chain Governance, DSRITB 2017, 845.

[MW17] Martini, M.; Weinzierl, Q.: Die Blockchain-Technologie und das Recht auf Vergessenwerden, NVwZ 2017, 1251.

[ST17] Schrey J.; Thalhofer T.: Rechtliche Aspekte der Blockchain, NJW 2017, 1431.

[Sz97] Szabo N.: Formalizing and Securing Relationships on Public Networks, first monday, Vol. 2, Nr. 9, 9/1997, https://ojphi.org/ojs/index.php/fm/article/view/548/469.

[37] Alle Internetquellen wurden zuletzt am 20.6.2019 überprüft.

Gestaltung smarter persönlicher Assistenten zwischen Rechtsverträglichkeit und Dienstleistungsqualität

Gestaltungsziele und Zielkonflikte

Robin Knote[1], Laura Friederike Thies[2], Matthias Söllner[3], Alexander Roßnagel[4], Jan Marco Leimeister[5]

Abstract: Smarte persönliche Assistenten von Amazon, Google und zahlreichen anderen Anbietern ermöglichen es, qualitativ hochwertige elektronische Dienstleistungen anzubieten. Gleichzeitig bringen diese Systeme auch viele Risiken mit sich. Berichte von diskriminierenden oder unverständlichem Systemverhalten häufen sich und verursachen Skepsis in der Gesellschaft. Diesen Problemen kann mit einer gleichermaßen rechtsverträglichen und qualitätsorientierten IT-Gestaltung entgegengewirkt werden. In diesem Beitrag werden Rechtsverträglichkeit und Dienstleistungsqualität als elementare Ziele für die Gestaltung smarter persönlicher Assistenten eingeführt. Zudem werden Zielkonflikte wie die Personalisierung von Leistungen bei gleichzeitiger Datensparsamkeit diskutiert und Implikationen für die Systemgestaltung erläutert.

Keywords/Stichwörter:
Smarte persönliche Assistenten, Rechtsverträglichkeit, Dienstleistungsqualität

1 Einleitung

Smarte persönliche Assistenten (SPA) befinden sich unter den prominentesten Anwendungen Künstlicher Intelligenz, die für den breiten Konsumentenmarkt verfügbar sind. Bis zum Jahr 2021, so eine Prognose, soll ihre Nutzerzahl weltweit auf 15,8 Millionen ansteigen [Tr16]. SPA wie Amazon Alexa oder Google Assistant nutzen Kontexterkennung und Kontextvorhersage, um ihren Nutzern je nach Ort, Zeit, Interessen und anderen Parametern passende Leistungen anzubieten [SS07, KS17]. Basis hierfür ist die Auswertung zahlreicher nutzer- und situationsrelevanter Daten, die über verschiedenen Sensoren wie Kameras, Mikrofone, Bewegungssensoren und über externe Datenquellen gesammelt werden. Über die passgenaue, auf Auswertung dieser Daten basierende Leistungserbringung hinaus, sind viele dieser Systeme auch in der Lage durch Wiederholungen

[1] University of Kassel, Information Systems, Pfannkuchstr. 1, 34121 Kassel, robin.knote@uni-kassel.de.
[2] University of Kassel, Public Law, Pfannkuchstr 1, 34121 Kassel, l.thies@uni-kassel.de.
[3] University of Kassel, Information Systems, Pfannkuchstr. 1, 34121 Kassel, soellner@uni-kassel.de.
[4] University of Kassel, Public Law, Pfannkuchstr. 1, 34121 Kassel, a.rossnagel@uni-kassel.de.
[5] University of Kassel, Information Systems, Pfannkuchstr. 1, 34121 Kassel, leimeister@uni-kassel.de.

doi:10.18420/inf2019_60

und Feedback zu lernen, also ihre Prognose- und Handlungsalgorithmen zu verbessern und so ihre Leistungen im Laufe der Zeit noch besser an die Bedürfnisse ihrer Nutzer anzupassen [RN10]. Außerdem profitieren die Nutzer von intuitiver Bedienbarkeit der SPA, die etwa mittels natürlicher Sprache steuerbar sind. Im Zusammenspiel dienen all diese Charakteristika dem Ziel den Nutzern elektronische Dienstleistungen in hoher Qualität zu bieten, was wichtig für die langfristige Zufriedenheit der Nutzer mit der Technik ist [Bl15, XBC13]. Dienstleistungsqualität meint dabei die Gesamtbewertung der Exzellenz von Dienstleistungsangeboten durch den Nutzer [Sa03]. Voraussetzung für eine hohe Dienstleistungsqualität von SPA ist die Verarbeitung einer großen Menge personenbezogener Daten. Je mehr brauchbare Daten dem System zur Verfügung gestellt werden, desto situationsadäquatere und personalisiertere Leistungen kann es für den jeweiligen Nutzer erbringen. Die Intransparenz bei der Sammlung, Verarbeitung und Weitergabe dieser Daten, wie auch die regelmäßig auftretenden Datenskandale zeichnen ein negatives Bild von Künstlicher Intelligenz und SPA im Besonderen und können die weitere Entwicklung und Kommerzialisierung dieser Systeme beeinträchtigen.

Aus rechtlicher Perspektive müssen Systeme, die personenbezogene Daten, also gemäß Art. 4 Nr. 1 Datenschutz-Grundverordnung (DSGVO) "alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person […] beziehen", im Einklang mit den entsprechenden rechtlichen Vorgaben gestaltet werden, um Risiken für den Einzelnen und für die Gesellschaft zu minimieren [Ho15]. Hier setzt das Konzept der Rechtsverträglichkeit an, das im Gegensatz zu Minimalansätzen, die lediglich eine rechtmäßige Gestaltung anstreben, die Vorgaben, die hochrangige grundrechtliche Normen aufstellen, in bestmöglicher Weise bei der Technikgestaltung umsetzen und so vor den negative Folgen der Techniknutzung bestmöglich Schutz bieten [HPR93a, HPR93b]. Viele Systementwicklungsprojekte betrachten rechtliche Anforderungen jedoch derzeit eher als "notwendiges Übel", das es im erlaubten Mindestmaß zu adressieren gilt.

Betrachtet man nun Rechtsverträglichkeit und Dienstleistungsqualität als gleichrangig wichtige Gestaltungsziele smarter persönlicher Assistenten, eröffnet sich ein weites Feld potentieller Zielkonflikte und konfligierender Anforderungen. So profitiert etwa die Dienstleistungsqualität von einer möglichst großen Menge personenbezogener Daten, während die DSGVO dem mit Art. 5 Grundsätze für die Verarbeitung personenbezogener Daten, wie Zweckbindung, Datenminimierung und Speicherbegrenzung entgegenstellt.

Infolgedessen sind die Anforderungsanalysten und Entwickler, die diese Systeme gestalten, regelmäßig mit komplexen Abwägungsentscheidungen zwischen Dienstleistungsqualität und Rechtsverträglichkeit konfrontiert. Hinzu kommt, dass insbesondere weniger erfahrene Entwickler sich oft nicht einmal bewusst sind, welche Dienstleistungsqualitäts- und Rechtsanforderungen für ein bestimmtes Problem oder Projekt von Relevanz sind.

Nach einer für die deutsche Bevölkerung repräsentativen Studie [EA18] liegen die häufigsten Gründe für die Nichtnutzung oder den Nichtkauf von SPA in Sicherheitsbeden-

ken wegen ungewollter Sprachaufnahmen (43 %), Angst vor massenhafter Sammlung personenbezogener Daten (39 %), mangelhafter Informationsqualität (35 %), Aversion gegen Kommunikation mit Maschinen (32 %), fehlenden Vorteilen durch Nutzung (30 %) und im Mangel an als nützlich empfundenen Funktionen (29 %).

Um diese Probleme zu adressieren und sie Lösungen zuzuführen, werden in diesem Beitrag Dienstleistungsqualität und Rechtsverträglichkeit, zwei sehr wichtige Gestaltungsziele für SPA, dar- und gegenübergestellt. Auf Grundlage der Operationalisierung dieser beiden Konzepte, entwickeln wir hochrangige Anforderungen für die Gestaltung von SPA, identifizieren potentielle Zielkonflikte und beschreiben Lösungsansätze. Dabei stellen wir sowohl problemorientiertes als auch lösungsorientiertes Wissen für Anforderungsanalysten und Systementwickler zur Verfügung. Damit wollen wir zu einer dienstleistungsorientierten und gleichzeitig rechtsverträglichen Gestaltung von SPA und anderen Anwendungen Künstlicher Intelligenz beitragen.

Dieser Beitrag ist wie folgt strukturiert: Nach dieser Einleitung werden in Abschnitt 2 die Grundlagen von SPA dargestellt. Anschließend werden in Abschnitt 3 als Gestaltungsziele für SPA betrachtet. Aufbauend darauf werden in Abschnitt 4 hochrangige Anforderungen beider Disziplinen vorgestellt und potentielle Konflikte identifiziert. Lösungsansätze werden dann in Abschnitt 5 diskutiert.

2 Smarte persönliche Assistenten

SPA, definiert als Systeme, “that use[s] input such as the user’s voice […] and contextual information to provide assistance by answering questions in natural language, making recommendations and performing actions” [Ba93, p. 223], unterstützen Nutzer bereits heute beim Einkaufen [Ve17], bei der Steuerung des Smart Home [Fe16], im Auto [Be14], oder als intelligenter Agent im Kundenservice [Xu17]. Frühere Forschungsarbeiten konnten fünf grundlegende Charakteristika von SPA identifizieren [Kn18].

- *Kontextsensitivität:* SPA bieten häufig die Möglichkeit der kontextabhängigen Personalisierung. Neben Daten von integrierten Sensoren, darunter auch Kameras und Mikrofone, können SPA Kontextinformationen auch von verschiedenen anderen Quellen beziehen, da sie in der Regel Teil eines größeren Sensornetzwerks sind (z.B. vernetzte Geräte im Smart Home).
- *Selbstlernfähigkeiten:* Zu einem gewissen Grad fungieren SPA als autonome, selbstlernende Agenten. Selbstlernfähigkeiten (in der Regel basierend auf maschinellem Lernen) finden sich in der Regel beim Sprachverständnis oder bei der Auswahl geeigneter Aktionen wieder. Amazon Alexa zum Beispiel "lernt", wie man aus undeutlichen Äußerungen durch lexikalische Approximation verwertbare Informationen herausfiltert.

- *Multimodalität:* SPA bieten Nutzern in der Regel verschiedene Arten der Interaktion an (z.B. per Sprache, Touchscreen oder mobile App). Je nach Vielfalt der Ein- und Ausgangskanäle lässt sich zwischen unidirektionaler und bidirektionaler Multimodalität unterscheiden.
- *Anthropomorphismus:* Anthropomorphismus meint "a conscious mechanism wherein people infer that a non-human entity has human-like characteristics and warrants human-like treatment" [Pu17, p. 2854]. Der Grad der „Vermenschlichung" unterscheidet sich je nach SPA, wird aber meist durch menschliche Sprachausgabe, virtuelle Charaktere oder eine Kombination aus beidem erreicht. [Pu17] untersuchte die Nutzerbewertungen von Alexa auf der Grundlage von vier verschiedenen Dimensionen: Personifizierungsgrad, Grad der Geselligkeit, Integration sowie technische Qualitäten und Probleme. Die Ergebnisse zeigen, dass eine stärkere Personifizierung durch den Nutzer mit häufigeren sozialen Interaktionen mit dem SPA einhergeht und dass Personifizierung so zu einem gewissen Grad das Nutzerverhalten bestimmen kann.
- *Plattformintegration und Erweiterbarkeit:* Ein SPA ist in der Regel Teil eines größeren Netzwerks der Dinge (z.B. Smart Home Geräte) sowie Teil der digitalen Infrastruktur eines Nutzers (z.B. Google Assistant als „Zugang" zum Google-Ökosystem). Auf diese Weise können SPA Daten aller Art aus entfernten Quellen verarbeiten, um auch umfassendere Benutzeranforderungen zu erfüllen.

3 Gestaltung smarter persönlicher Assistenten zwischen Rechtsverträglichkeit und Dienstleistungsqualität

Es lässt sich vermuten, dass rechtsverträgliche und dienstleistungsorientierte SPA vor dem Hintergrund der vorherrschenden Entwicklungspraxis auf große Akzeptanz in der Gesellschaft stoßen werden. Bei vielen Systementwicklungsprojekten wird rechtlichen Anforderungen nur wenig Aufmerksamkeit gewidmet. Die daraus resultierenden Produkte erfüllen dann häufig die jeweiligen rechtlichen Vorgaben nur im absoluten Mindestmaß. Solche Ansätze können Techniknutzer nicht ausreichend vor sozialen und individuellen Risiken bewahren; außerdem besteht ein Risiko für Datenpannen und nicht zuletzt können im Falle von Gesetzesänderungen horrende Kosten für nachträgliche Systemänderungen, Imagebereinigung und Kundenrückgewinnung anfallen [Ho15]. Das Konzept der Rechtsverträglichkeit hingegen verfolgt das Ziel, Techniknutzer langfristig bestmöglich vor den Risiken der Technik zu schützen [Ho15]. Um Technik rechtsverträglich zu gestalten, werden dazu die relevanten (grund-)rechtlichen Vorgaben identifiziert, aus denen dann in mehreren Schritten konkrete technische Gestaltungsvorschläge abgeleitet werden [HPR93b]. Dabei enthalten viele gesetzliche Normen unbestimmte Rechtsbegriffe, die zu ihrer Anwendung einer näheren, durch Auslegung zu ermittelnden Bestimmung bedürfen [Sc18, § 40, para. 152 et seq.] und sich demnach nicht ohne Weiteres unmittelbar in technische Gestaltungsziele übersetzen lassen. So schreibt etwa Art.

25 Abs. 1 DSGVO dem für die Datenverarbeitung Verantwortlichen vor, unter Berücksichtigung mehrerer Abwägungskriterien wie „Stand[…] der Technik“, „Implementierungskosten“ und „Eintrittswahrscheinlichkeit und Schwere der mit der Verarbeitung verbundenen Risiken für die Rechte und Freiheiten natürlicher Personen“ „geeignete technische und organisatorische Maßnahmen“ zu treffen, um den Datenschutzgrundsätzen durch Technikgestaltung und datenschutzfreundliche Voreinstellungen zur Durchsetzung zu verhelfen. Und obwohl die “Pseudonymisierung” als ein Beispiel für eine solche technische Maßnahme genannt wird, bleibt dem Verantwortlichen insgesamt ein weiter Gestaltungsspielraum [Kü18, Art. 25 DSGVO, para. 16 f.]. So könnte der Verantwortliche den Grundsatz der Datenminimierung aus Art. 5 Abs. 1 lit. e) DSGVO, der im Wesentlichen besagt, dass die Datenverarbeitung im Rahmen der Zweckbindung “qualitativ und quantitativ” begrenzt werden muss [Fr18], auf verschiedene Weisen umsetzen. Er könnte die Datenverarbeitung so gestalten, dass unter Angabe eines weit gefassten Zwecks die Speicherung vieler Daten in vielen Verarbeitungsschritten zur Zweckerreichung erforderlich ist. Im Gegensatz zu einem solchen, lediglich auf die Untergrenze der Rechtmäßigkeit zielenden Ansatz, könnte der Verantwortliche, um die rechtliche Vorgabe besser umzusetzen, auch unter Angabe eines eng gefassten Zweckes wenige Daten in wenigen Verarbeitungsschritten verarbeiten und sie anschließend endgültig löschen. Neben den Vorteilen, die ein solcher Ansatz für die Grundrechtsverwirklichung der Nutzer bietet, kann die rechtsverträgliche Gestaltung – insbesondere zu Zeiten regelmäßig auftretender Datenskandale – von KI-Systemen auch kompetitive Vorteile gegenüber internationalen Wettbewerbsprodukten bringen [Al13].

Die Dienstleistungsqualität, also die Qualität der Leistungserbringung, ist ein wesentlicher Faktor für die Zufriedenheit der Nutzer mit einem System und damit entscheidend für den wirtschaftlichen Erfolg [Br16, DM03]. Das Konzept der Dienstleistungsqualität beschreibt die Fähigkeit eines Anbieters oder eines an dessen Stelle auftretenden Systems, den Kundenerwartungen bei der Leistungserbringung auf einem bestimmten Anforderungsniveau gerecht zu werden [Br16]. Die Differenz zwischen erwarteten und tatsächlich wahrgenommenen Faktoren ist dabei Element des kontinuierlichen Spektrums aller möglichen Anforderungsniveaus. In Bezug auf die Systementwicklung lassen sich jedoch – analog zur Rechtsverträglichkeit – zwei wesentliche Ausprägungen beschreiben. Denn viele Systementwicklungsprojekte fokussieren auf die Implementierung der funktionalen Aspekte eines Systems nach Lastenheft oder Backlog und schenken wissenschaftlich fundierte Faktoren der Qualitätswahrnehmung von Nutzern weniger Aufmerksamkeit. Dies kann dazu führen, dass, obgleich, objektiv betrachtet, womöglich ein großer Nutzen durch die Systemnutzung entstünde, dieser von Nutzern nicht in vollem Maße erkannt und genutzt werden kann, weil die minimalistische Qualitätsorientierung zu erhöhten Absprungraten in frühen Nutzungsphasen führt oder bereits vorab bewusst die Entscheidung getroffen wird, ein qualitativ höherwertig erscheinendes Konkurrenzsystem zu nutzen. So könnte beispielsweise ein Nutzer einen SPA, der auf Basis der eigenen Präferenzen in der Lage ist, personalisierte Dienste, wie das Abspielen des Lieblingsliedes, auszuführen, einem SPA vorziehen, der eine solche Personalisierung nicht zulässt. Denn Personalisierung der Funktionalitäten ist dabei ein Faktor für Dienst-

leistungsqualität und somit für Nutzer ein Indikator, wie zufrieden sie mit dem SPA während der Nutzung voraussichtlich sein werden [LH11].

Betrachtet man nun diese beiden Perspektiven in der Gegenüberstellung, ergibt sich ein Lösungsraum für die Gestaltung von SPA. Idealerweise werden dabei sowohl die Rechts, als auch die Qualitätsziele in vollem Umfang verwirklicht. Trotz dieses Bestrebens allen Zielen gerecht zu werden, können bei der Systementwicklung zahlreiche Zielkonflikte auftreten, mit denen Anforderungsanalysten und Entwickler konfrontiert werden. Sofern nicht beide Ziele vollumfänglich verwirklicht werden können, können die Lösungen entweder verstärkt rechtsorientiert oder verstärk dienstleistungsorientiert aussehen. Pauschal lässt sich keine dieser beiden Gestaltungsalternativen bevorzugt vor der anderen empfehlen. Eine Firma, die eine Innovation auf den Markt bringen will, könnte, weil sie von einer raschen Markteinführung profitieren will, die Dienstleistungsqualität der Rechtsverträglichkeit gegenüber priorisieren. Gleichwohl ist zu berücksichtigen, dass bei einer schnellen Anhäufung vieler personenbezogener Daten die nachträgliche Implementierung von Rechtszielen eine enorme betriebsökonomische Belastung darstellen kann. Eine Kompromisslösung aus den Minimalanforderungen beider Welten sollte, wenn möglich, vermieden werden. Es lässt sich beobachten, dass insbesondere die aktuellen kommerziellen SPA sich in diesem Bereich bewegen. In jedem Fall zu vermeiden sind Umsetzungen, die rechtswidrig und/oder qualitativ minderwertig sind. Diese Lösungen sind entweder wenig nützlich und/oder verstoßen gegen geltendes Recht

4 Gestaltungsziele für smarte persönliche Assistenten

4.1 Rechtliche Ziele

Um rechtsverträglich zu sein, müssen SPA im europäischen Rechtsraum unter anderem mit den Artikeln 7, 11, 16 und insbesondere mit Artikel 8 der Europäischen Grundrechtecharta, der den Schutz personenbezogener Daten garantiert [St16], vereinbar sein. Die DSGVO, die im Mai vergangenen Jahres Geltung erlangte, ist eine sekundärrechtliche Konkretisierung des Art. 8 der Europäischen Grundrechtecharta. Sie trat an Stelle des bis dahin geltenden Bundesdatenschutzgesetzes. Als Verordnung ist sie unmittelbar in den Mitgliedstaaten anwendbar und bedarf keiner nationalen Umsetzungsgesetze (EuGH, Urt. v. 10. Oktober 1973 – C-34/73, ECLI:EU:C:1973:101 Rn. 10 – Variola; EuGH, Urt. v. 9. März 1978 – C-106/77, ECLI:EU:C:1978:49 Rn. 14 – Simmenthal II). Aufgrund des Anwendungsvorrangs des Europäischen Rechts geht das unionale Recht im Konfliktfall dem nationalen Recht vor (EuGH, Urt. v. 15. Juli 1964 – C-6/64, ECLI:EU:C:1964:66 Rn. 3 – Costa/ENEL; EuGH, Urt. v. 17. Dezember 1970 – C-11/70, ECLI:EU:C:1970:114 Rn. 3 – Internationale Handelsgesellschaft; EuGH, Urt. v. 9. März 1978 – C-106/77, ECLI:EU:C:1978:49 Rn. 17 f. – Simmenthal II). Nationales Recht kann neben der DSGVO nur zur Anwendung kommen, wenn es das europäische Recht präzisiert, konkretisiert oder Öffnungsklauseln aus der Verordnung dem nationalen Gesetzgeber ermöglichen Regelungen zu erlassen [Ro18, § 2, Rn. 15 ff.].

Gemäß Art. 3 Absatz 1 und 2 DSGVO, findet dieser sekundärrechtliche Rechtsakt "Anwendung auf die Verarbeitung personenbezogener Daten, soweit diese im Rahmen der Tätigkeiten einer Niederlassung eines Verantwortlichen oder eines Auftragsverarbeiters in der Union erfolgt, unabhängig davon, ob die Verarbeitung in der Union stattfindet" sowie auch auf von nicht in der Union niedergelassenen Verantwortlichen, wenn Personen aus der Union involviert sind Die DSGVO normiert zahlreiche Vorgaben für die Verarbeitung personenbezogener Daten. Zunächst bedarf, gemäß Art. 6 DSGVO, jede Datenverarbeitung einer rechtlichen Grundlage. Für SPA, die oft gekoppelt mit einem Vertrag über Service-Dienstleistungen angeboten werden, stellt Art. 6 lit. b) DSGVO regelmäßig die Rechtsgrundlage der Datenverarbeitung dar. Eine Datenverarbeitung nach Art. 6 lit. b) ist rechtmäßig, "wenn die Verarbeitung für die Erfüllung eines Vertrags, dessen Vertragspartei die betroffene Person ist, oder zur Durchführung vorvertraglicher Maßnahmen erforderlich, die auf Anfrage der betroffenen Person erfolgen." Solche gekoppelten Verträge können jedoch gegen Art. 7 Unterabsatz 4 DSGVO verstoßen. Denn laut Artikel 7 Unterabsatz 4 DSGVO muss bei der Beurteilung, ob eine Einwilligung in eine Datenverarbeitung freiwillig erteilt wurde, auch berücksichtigt werden muss, ob die Erfüllung des Vertrages von der Einwilligung in die Verarbeitung personenbezogener Daten abhängig gemacht wird, obwohl diese Daten nicht für die Vertragserfüllung erforderlich sind.

In Bezug auf SPA kommt neben Art. 6 lit. b) DSGVO als Grundlage für die Verarbeitung personenbezogener Daten noch die Einwilligung der betroffenen Person in die Datenverarbeitung nach Art. 6 lit. a) DSGVO in Betracht. Die Voraussetzungen, unter denen eine solche Einwilligung rechtmäßig ist, lassen sich den Art. 4 Nr. 11, 6 Abs. 1 lit. a) und 7 DSGVO in Verbindung mit den Erläuterungen entnehmen [Kü18]. An dieser Stelle ist zu berücksichtigen, dass gegen Treu und Glauben verstößt, wenn der Verantwortliche für eine Datenverarbeitung, die schon auf Art. 5 Abs. 1 lit. b) DSGVO gestützt wird, zusätzlich noch eine Einwilligung einholt. Denn der nach Art. 7 Abs. 3 S. 3 DSGVO geforderte Hinweis auf ein Widerspruchsrecht führte in diesem Fall in die Irre.

Daneben müssen bei jeder Datenverarbeitung die Grundsätze für die Verarbeitung personenbezogener Daten aus Art. 5 DSGVO, namentlich Rechtmäßigkeit, Verarbeitung nach Treu und Glauben, Transparenz, Zweckbindung, Datenminimierung, Richtigkeit, Speicherbegrenzung und Integrität und Vertraulichkeit, eingehalten werden [Kü18, Art. 5, Rn. 1 ff]. Bei SPA, die häufig in privaten Kontexten verwendet werden, können Daten, aus denen "die rassische und ethische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen", "die Gewerkschaftszugehörigkeit" oder "genetische[…] Daten, biometrische[…] Daten", "Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung" anfallen. Diese besonderen Kategorien personenbezogener Daten unterliegen den strengeren Vorgaben des Art. 9 DSGVO.

Im Anwendungsbereich des Europäischen Rechts gelten auch die europäischen Grundrechte. Ausgehend von den für die Nutzer von SPA relevanten Grundrechten und einer Analyse der mit der Technik einhergehenden Chancen und Risiken lassen sich aus rechtlicher Perspektive folgende Gestaltungsziele identifieren (Tabelle 1).

Transparenz	Authentifikation
Zweckbindung	Kontrollierbarkeit
Datenminimierung	Schutz der Privatsphäre
Speicherbegrenzung	Verhältnismäßigkeit
Integrität	Keine Diskriminierung
Vertraulichkeit	

Tab. 1: Gestaltungsziele zur Steigerung der Rechtsverträglichkeit von SPA

4.2 Dienstleistungsqualitätsziele

Als Gütemaß für Dienstleistungen kristallisierte sich seit Mitte der 80er Jahre die Dienstleistungsqualität (engl. Service Quality) aus dem Marketing heraus [PZB85]. Zunächst primär für die Analyse der Interaktion zwischen Menschen im Dienstleistungsprozess konzipiert, wurden auf Basis erster Erhebungsinstrumente der Dienstleistungsqualität später Methoden und Techniken entwickelt, deren Fokus auf spezielleren Anwendungsbereichen liegt. Der für SPA relevante wahrgenommene Nutzen und die Bereitschaft zur langfristigen, intensiven Nutzung von Technologie hängen stark von der Dienstleistungsqualität ab [VB08]. Die Forschung um Dienstleistungsqualität sucht primär nach Antworten auf die Frage nach den Erwartungen der Nutzer und wie diese erfüllt werden können [PZB85, PZB88]. Insbesondere für elektronische Dienste (E-Services), die sich Anfang der 2000er Jahre durchgesetzt haben, existiert heute ein reicher Fundus an Erfolgsfaktorenforschung [17, 25]. Nach Wissen der Autoren existiert für den besonderen Fall der SPA jedoch noch keine eigenständige Erfolgsfaktorenforschung, was eine Untersuchung und ggf. Modifikation bereits vorhandener, ähnliche Dienstleistungsarten adressierender Modelle notwendig macht. SPA können dabei als eine, von den von [Kn18] identifizierten Merkmalen geprägte Form sog. Self-Service-Technologien betrachtet werden. Nutzer von Self-service-Technologien erbringen die Leistung eigenständig in direkter Interaktion mit dem smarten Objekt, während der eigentliche Dienstleistungsanbieter in den Hintergrund tritt. Für diese Art von Dienstleistungen existieren bereits etablierte Theorien. [LH11] beschreibt sieben Qualitätsmerkmale von Self-Service-Technologien: nämlich Funktionalität, Freude, Sicherheit/Datenschutz, Zuverlässigkeit, Ästhetik, Komfort und Anpassbarkeit.

Um die Dienstleistungsqualität von SPA so vollständig und präzise wie möglich abzubilden, sollten die vorgenannten Qualitätsfaktoren an die spezifischen SPA-Eigenschaften und die aktuelle Forschung angepasst werden. Beispielsweise zeigt die Forschung zu SPA, dass das Vertrauen in das System eine wichtige Rolle für Nutzer spielt [SBA17]. Auch die durch Anthropomorphismus erreichte soziale Präsenz und Personalisierung sind wichtige Schlüsselfaktoren in der Interaktion mit SPA [GMM17].

Basierend auf theoretischen Grundlagen, empirischen Erkenntnissen und eigenen Konzeptualisierungen lassen sich insgesamt 11 Gestaltungsziele zur Steigerung der Dienstleistungsqualität von SPA ableiten (Tabelle 2).

Funktionalität	Soziale Präsenz
Freude	Vertrauen
Zuverlässigkeit	Empathie
Ästhetik	Informationsgehalt
Komfort	Kontinuierliche Entwicklung
Personalisierung	

Tab. 2: Gestaltungsziele zur Steigerung der Dienstleistungsqualität von SPA

4.3 Potenzielle Zielkonflikte

Bei der Spezifikation von SPA kann es zwischen den Zielen der Dienstleistungsqualität und der Rechtsverträglichkeit zu mehreren Zielkonflikten kommen. Nachfolgend sind solche Zielkonflikte gruppiert nach den fünf grundlegenden Eigenschaften von SPA exemplarisch dargestellt:

1. *Kontextsensitivität* ist eine Grundvoraussetzung für die Bereitstellung personalisierter Dienste. Daher werden personenbezogene Daten erhoben und verarbeitet. Rechtsverträglichkeit erfordert jedoch bestenfalls die Vermeidung von Daten mit persönlichem Bezug oder zumindest die Reduzierung und schnelle Löschung. Da sich der Personenbezug von Daten negativ auf die Rechtsverträglichkeit, jedoch durch die Personalisierung positiv auf die Dienstleistungsqualität auswirkt, verkörpert dieser Konflikt den Kerngedanken des sog Personalisierungs-Privatsphäre-Paradoxons, welches das Verhalten von Nutzern in einer solchen Konfliktsituation beschreibt und bereits für andere Zusammenhänge (bspw. personalisierte Werbung) ausgiebig beforscht ist [Ka17, GZS16, Su13, AK06]. Darüber hinaus bedarf die Verarbeitung personenbezogener Daten gemäß Artikel 6 DSGVO einer Rechtsgrundlage, wie beispielsweise der Einwilligung des Nutzers. Die typische Länge von Datenschutzvereinbarungen überfordert Nutzer jedoch häufig, sodass wenige dazu bereit sind, sie in Gänze zu lesen [JPJ05]. Hinzu kommen häufig schwammige Formulierungen zu den Verwendungszwecken der personenbezogenen Daten. Rechtsverträgliche Gestaltung von SPA zielt jedoch darauf ab, den Nutzer über die Einzelheiten der Verarbeitung personenbezogener Daten in einfach verständlicher Weise zu informieren.

2. *Selbstlernfähigkeiten* beruhen auf der (in der Regel algorithmischen) Verarbeitung personenbezogener Daten, um die Dienstleistungsqualität und damit das Erlebnis für den Nutzer im Laufe der Zeit zu verbessern. Daher können für dieses Ziel ähnliche Konflikte auftreten wie für die Kontextsensitivität. Obwohl [Su13] jedoch nahelegt, dass die rein lokale (d.h. on-device) Verarbeitung personenbezogener Daten für Personalisierungszwecke aus datenschutzrechtlicher Sicht unbedenklicher ist als etwa die Verarbeitung in der Cloud, ist dies bei den meisten kommer-

ziellen SPA nicht vorgesehen. Aufgrund der hohen Mengen heterogener Daten, die zusammengefasst und ausgewertet werden müssen, ist eine kontinuierliche, zielgerichtete und sich jederzeit an den individuellen Wünschen und Vorlieben der Nutzer orientierende Selbstlernfähigkeit technisch umständlich zu realisieren. Selbst bei Gelingen einer solchen Umsetzung würde aus rechtlicher Sicht die Frage aufkommen, inwieweit eine solch autonome Erhebung und Verarbeitung personenbezogener Daten für Selbstlernzwecke von SPA den Zielen der Rechtsverträglichkeit entspricht.

3. *Anthropomorphismus* zielt darauf ab, soziale Beziehungen mit dem SPA zu etablieren [Pu17]. Obwohl diese Eigenschaft sich nicht direkt auf die Erhebung und Verarbeitung personenbezogener Daten bezieht, werden menschähnliche Merkmale wie empathisches Verhalten als sozial wirksam angesehen. Eine nicht-triviale offene Frage, sowohl aus rechtlicher als auch aus qualitativer Sicht, ist dabei, inwieweit menschliche Attributionen und soziale Bindungen die Bereitschaft des Nutzers beeinflussen, dem SPA personenbezogene Daten zu offenbaren. Anthropomorphismus kann zudem auch zu mehr Transparenz führen, indem das Systemverhalten einfühlsam und leicht verständlich erklärt wird.

4. *Multimodalität* bedeutet, verschiedene Interaktionskanäle bereitzustellen, um sowohl die Benutzerfreundlichkeit als auch Freude und Komfort im Umgang mit dem SPA zu verbessern. Jeder dieser Kanäle ist jedoch auch ein potenzieller Angriffspunkt, über den unberechtigte Dritte in das System eindringen und personenbezogene Daten ausspähen können. Alle Interaktionskanäle sollten daher eigenständig durch adäquate technische Maßnahmen geschützt und der Zugang zum System auf eine ausgewählte Gruppe von Personen beschränkt werden.

5. *Plattformintegration und Erweiterbarkeit* meint die Verbindung von physischen Objekten und digitaler Infrastruktur mit dem SPA. Durch die Erweiterung der digitalen und physischen Infrastruktur um den SPA entstehen ebenfalls potenzielle Angriffspunkte für unberechtigte Dritte, um personenbezogene Daten auszuspähen. Diese Gefahr wird jedoch bei den meisten kommerziellen SPA zugunsten erhöhten Komforts und Benutzerfreundlichkeit billigend in Kauf genommen. Beispielsweise ist Amazons Alexa mit den Amazon Web Services verbunden, die auch als Infrastruktur für die Amazon-Shopping-Seite dienen. Willigt der Nutzer entsprechend ein, erhält der SPA Zugriff auf das persönliche Profil des Nutzers sowie seine Einkaufshistorie auf Basis dessen der SPA lernen und dem Nutzer personalisierte Angebote machen kann. Aufgrund seiner Einbettung ist ein SPA auch ein weiterer Zugangspunkt zu personenbezogenen Daten über die verbundenen Geräte. Daher ist es wichtig, dass der Benutzer jederzeit Kontrolle und Transparenz über die Datenströme innerhalb der vernetzten Infrastrukturen behält.

5 Diskussion und Zusammenfassung

In diesem Beitrag haben wir Rechtsverträglichkeit und Dienstleistungsqualität als wichtige Gestaltungsziele für SPA vorgestellt. Zwischen beiden Gestaltungsbereichen gibt es zahlreiche potenzielle Zielkonflikte, aus denen in diesem Beitrag lediglich eine Auswahl exemplarisch diskutiert werden kann. In einem nächsten Forschungsschritt sollen auf Basis der Ziele entsprechende Anforderungen an SPA anhand spezifischer Anwendungsfälle identifiziert werden. Um eine solche Operationalisierung zu gewährleisten, eignen sich insb. zielorientierte Ansätze des Anforderungsmanagements (Goal-oriented requirements engineering, GORE) [va01, va13]. Dadurch lassen sich zu Zielen entsprechende Unterziele und Systemanforderungen spezifizieren und miteinander in Verbindung setzen, wodurch zum einen ein vollständiges Bild der Anforderungslandschaft erstellt und zum anderen weitere Zielkonflikte aufgedeckt und beschrieben werden können. Anschließend sollen entsprechende Lösungsalternativen für diese Konflikte in interdisziplinären Workshops mit Dienstleistungs-, Rechts- und Entwicklungsexperten gesammelt und bewertet werden.

Um die Wiederverwendbarkeit und die praktische Anwendbarkeit des so entstehenden Problem- und Lösungswissens für Systementwicklungsprozesse zugänglich zu machen, sollte es in strukturierter und Entwicklungspraktikern bekannter Form dokumentiert werden. Dafür eignen sich insbesondere sog. Muster, wiederverwendbare Vorlagen, die wiederkehrende Probleme spezifizieren und den Kern ihrer Lösung beschreiben [Al79, Ga94, WF11]. Generell lässt sich zwischen zwei Arten von Mustern unterscheiden: Anforderungsmuster helfen dem Anforderungsanalysten (bzw. Product Owner) dabei, projektspezifische Systemanforderungen auf Basis von Zielen der Rechtsverträglichkeit und der Dienstleistungsqualität festzulegen [Fr10]. Entwurfsmuster helfen Systementwicklern für eine Reihe von Anforderungen passende Gestaltungselemente auszuwählen und diese so zu implementieren, dass die Gestaltungsziele bestmöglich erreicht werden [Ba14]. Frühere Untersuchungen haben gezeigt, dass die Verwendung von Mustern positive Auswirkungen auf die Entwicklungseffizienz, die Zusammenarbeit zwischen den Beteiligten und die Qualität von Anforderungsspezifikationen und zu entwickelnden Systemen haben kann[Ho14]. Zusätzlich liefern Anforderungs- und Entwurfsmuster im Kontext von SPA das nötige Wissen, um bereits vor der Entwicklung eines Systems die beschriebenen Zielkonflikte systematisch aufzulösen und somit auf effizientere Weise zielkonforme SPA zu entwickeln.

6 Literaturverzeichnis

[AK06] Awad, N. F.; Krishnan, M. S.: The personalization privacy paradox. An empirical evaluation of information transparency and the willingness to be profiled online for personalization. In Mis Quarterly, 2006, 30; pp. 13–28.

[Al13] Albrecht, J. P.: Starker EU-Datenschutz wäre Standortvorteil. In Datenschutz und Datensicherheit - DuD, 2013, 37; pp. 655–657.

[Al79] Alexander, C.: The timeless way of building. Oxford University Press, New York, 1979.

[Ba14] Baraki, H. et al.: Towards Interdisciplinary Design Patterns for Ubiquitous Computing Applications. Kassel University Press, Kassel, 2014.

[Ba93] Baber, C.: Developing interactive speech technology. Taylor & Francis, Inc, 1993.

[Be14] Bengler, K. et al.: Three Decades of Driver Assistance Systems. Review and Future Perspectives. In IEEE Intelligent Transportation Systems Magazine, 2014, 6; pp. 6–22.

[Bl15] Blut, M. et al.: E-Service Quality. A Meta-Analytic Review. In Journal of Retailing, 2015, 91; pp. 679–700.

[Br16] Bruhn, M.: Qualitätsmanagement für Dienstleistungen. Handbuch für ein erfolgreiches Qualitätsmanagement. Grundlagen Konzepte Methoden. Springer Berlin Heidelberg; Imprint: Springer Gabler, Berlin, Heidelberg, 2016.

[DM03] DeLone, W. H.; McLean, E. R.: The DeLone and McLean Model of Information Systems Success: A Ten-Year Update. In Journal of Management Information Systems, 2003, 19; pp. 9–30.

[EA18] EARSandEYES GmbH: Welche Gründe sprechen für Sie gegen eine Nutzung von Sprachassistenten? https://de.statista.com/statistik/daten/studie/872316/umfrage/gruende-fuer-die-nichtnutzung-von-sprachassistenten-in-deutschland/, accessed 26 Nov 2018.

[Fe16] Fernando, N. et al.: Examining Digital Assisted Living: Towards a Case Study of Smart Homes for the Elderly: Proceedings of the 24th European Conference on Information Systems (ECIS), Istanbul, Turkey, 2016.

[Fr10] Franch, X. et al.: A Metamodel for Software Requirement Patterns. In (Wieringa, R.; Persson, A. Eds.): Requirements Engineering: Foundation for Software Quality. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010; pp. 85–90.

[Fr18] Frenzel, E. M.: Art. 5 Grundsätze für die Verarbeitung personenbezogener Daten. In (Paal, B. P.; Ernst, S.; Pauly, D. A. Eds.): Datenschutz-Grundverordnung Bundesdatenschutzgesetz. C.H. Beck, München, 2018; pp. 65–85.

[Ga94] Gamma, E. et al.: Design Patterns - Elements of Reusable Object-Oriented Software. Addison-Wesley, Reading, 1994.

[GMM17] Gnewuch, U.; Morana, S.; Maedche, A.: Towards Designing Cooperative and Social Conversational Agents for Customer Service. In Thirty Eighth International Conference on Information Systems (ICIS 2017), 2017.

[GZS16] Guo, X.; Zhang, X.; Sun, Y.: The privacy-personalization paradox in mHealth services acceptance of different age groups. In Electronic Commerce Research and Applications, 2016, 16; pp. 55–65.

[Ho14] Hoffmann, A.: Anforderungsmuster zur Spezifikation soziotechnischer Systeme. Standardisierte Anforderungen der Vertrauenwürdigkeit und Rechtsverträglichkeit. Kassel University Press, Kassel, Germany, 2014.

[Ho15] Hoffmann, A. et al.: Legal Compatibility as a Characteristic of Sociotechnical Systems. In Business & Information Systems Engineering, 2015, 57; pp. 103–113.

[HPR93a] Hammer, V.; Pordesch, U.; Roßnagel, A.: KORA-eine Methode zur Konkretisierung rechtlicher Anforderungen zu technischen Gestaltungsvorschlägen für Informations- und Kommunikationssysteme. In Infotech/I+ G, 1993, 21.

[HPR93b] Hammer, V.; Pordesch, U.; Roßnagel, A.: Betriebliche Telefon- und ISDN-Anlagen rechtsgemäß gestaltet. Universitätsbibliothek Kassel, Kassel, 1993.

[JPJ05] Jensen, C.; Potts, C.; Jensen, C.: Privacy Practice of Internet Users: self-reports versus observed behavior. In International Journal of Human-Computer Studies, 2005, 63; pp. 203–227.

[Ka17] Karwatzki, S. et al.: Beyond the Personalization–Privacy Paradox. Privacy Valuation, Transparency Features, and Service Personalization. In Journal of Management Information Systems, 2017, 34; pp. 369–400.

[Kn18] Knote, R. et al.: The What and How of Smart Personal Assistants: Principles and Application Domains for IS Research. In Multikonferenz Wirtschaftsinformatik (MKWI), 2018.

[KS17] Knote, R.; Söllner, M.: Towards Design Excellence for Context-Aware Services - The Case of Mobile Navigation Apps: 13th International Conference on Wirtschaftsinformatik (WI), St. Gallen, Switzerland, 2017.

[Kü18] Kühling, J. et al.: Datenschutz-Grundverordnung. C.H. Beck, München, 2018.

[LH11] Lin, J.-S. C.; Hsieh, P.-L.: Assessing the Self-service Technology Encounters: Development and Validation of SSTQUAL Scale. In Journal of Retailing, 2011, 87; pp. 194–206.

[Pu17] Purington, A. et al.: "Alexa is my new BFF": Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems. ACM Press, New York, New York, USA, 2017; pp. 2853–2859.

[PZB85] Parasuraman, A.; Zeithaml, V. A.; Berry, L. L.: A Coneptual Model of Service Quality and its Implilcations for Future Research. In Journal of Marketing, 1985, 49; pp. 41–50.

[PZB88] Parasuraman, A.; Zeithaml, V. A.; Berry, L. L.: SERVQUAL: A Multiple-Item Scale for Measuring Customer Perceptions of Service Quality. In Journal of Retailing, 1988, 64.

[RN10] Russell, S. J.; Norvig, P.: Artificial intelligence. A modern approach. Prentice Hall, Upper Saddle River, N.J., 2010.

[Ro18] Roßnagel, A. et al.: Das neue Datenschutzrecht Europäische Datenschutz-Grundverordnung und deutsche Datenschutzgesetze. Nomos, Baden-Baden, 2018.

[Sa03] Santos, J.: E-service quality: a model of virtual service quality dimensions. In Managing Service Quality: An International Journal, 2003, 13; pp. 233–246.

[SBA17] Saffarizadeh, K.; Boodraj, M.; Alashoor, T. M.: Conversational Assistants: Investigating Privacy Concerns, Trust, and Self-Disclosure. In Thirty Eighth International Conference on Information Systems (ICIS 2017), 2017.

[Sc18] Schmitz, H. et al.: Verwaltungsverfahrensgesetz Kommentar. C.H. Beck, München, 2018.

[SS07] Sitou, W.; Spanfelner, B.: Towards Requirements Engineering for Context Adaptive Systems: 31st Annual International Computer Software and Applications Conference, 2007. IEEE Computer Society, Los Alamitos, CA, USA, 2007; pp. 593–600.

[St16] Stern, K. et al.: Europäische Grundrechte-Charta GRCh Kommentar. C.H. Beck, München, 2016.

[Su13] Sutanto, J. et al.: Addressing the Personalization-Privacy Paradox: An Empirical Assessment from a Field Experiment on Smartphone Users. In Mis Quarterly, 2013, 37; pp. 1141–1164.

[Tr16] Tractica: The Virtual Digital Assistant Market Will Reach $15.8 Billion Worldwide by 2021. https://www.tractica.com/newsroom/press-releases/the-virtual-digital-assistant-market-will-reach-15-8-billion-worldwide-by-2021/, accessed 20 Aug 2017.

[va01] van Lamsweerde, A.: Goal-oriented requirements engineering: a guided tour: Proceedings. Fifth IEEE International Symposium on Requirements Engineering August 27-31, 2001, Royal York Hotel, Toronto, Canada. IEEE Computer Society, Los Alamitos, Calif, 2001; pp. 249–262.

[va13] van Lamsweerde, A.: Requirements engineering. From system goals to UML models to software specifications. Wiley, Chichester [u.a.], 2013.

[VB08] Venkatesh, V.; Bala, H.: Technology Acceptance Model 3 and a Research Agenda on Interventions. In Decision Sciences, 2008, 39; pp. 273–315.

[Ve17] Venkatesh, V. et al.: Design and Evaluation of Auto-ID Enabled Shopping Assistance Artifacts in Customer's Mobile Phones: Two Retail Store Laboratory Experiments. In Mis Quarterly, 2017, 41; pp. 83–113.

[WF11] Wellhausen, T.; Fießer, A.: How to write a pattern? A rough guide for first-time pattern authors. http://europlop.net/sites/default/files/files/0_How%20to%20write%20a%20pattern-2011-11-30_linked.pdf, accessed 22 Feb 2016.

[XBC13] Xu, J.; Benbasat, I.; Cenfetelli, R. T.: Integrating Service Quality with System and Information Quality. An Empirical Test in the E-Service Context. In Mis Quarterly, 2013, 37; pp. 777–794.

[Xu17] Xu, A. et al.: A New Chatbot for Customer Service on Social Media: Proceedings of the Annual CHI Conference on Human Factors in Computing Systems. ACM, New York, NY, 2017; pp. 3506–3510.

Extended Abstracts

Eine Übersicht und Vergleich zur NIS-Richtlinie innerhalb der EU mit dem Fokus auf die Betreiber von Kritischen Infrastrukturen

Philipp Schütz[1]

Keywords: NIS-Richtlinie; Vergleich-NIS-Richtlinie; KRITIS-Schutz; KRITIS-Betreiber

1 Einleitung

Das vorliegende extended Abstract fasst auszugsweise die Erkenntnisse der Veröffentlichung [SC19] zum 16. Deutschen IT-Sicherheitskongress beim Bundesamt für Sicherheit in Informationstechnik vom 21. Bis 23. Mai 2019 in Bonn-Bad Godesberg zusammen. Der vollständige Artikel ist zum Kongresstag im Kongressband erschienen.

Die europäische Gesetzgebung hat mit der NIS-Richtlinie als legislativem Rahmen erstmalig dezidierte Informationssicherheitsanforderungen und -ziele zum Schutz von Kritischen Infrastrukturen und deren übergreifende staatlichen Kollaborationsstrukturen verabschiedet.

2 Erkenntnisinteresse und Ziele

Die zentralen sechs Fragestellungen sind:

1. Ist die NIS-RL in nationales Recht umgesetzt, und wenn ja, über welche nationalen Hoheitsakte (Gesetze, Verordnungen, Dekrete und Cyber-Sicherheitsstrategien)?
2. Existieren ausgeprägte KRITIS-Sektoren, Subsektoren und werden Arten von Betreibern genannt (wie bspw. Stromverteilnetzbetreiber, Wasserversorger)?
3. Über welche Mechanismen findet eine Betreiber-Identifikation statt (etwa analog zu Deutschland über öffentlich zugängliche quantitative und qualitative Schwellwerte, wie in den KRITIS-Verordnungen)?
4. Im Falle von gravierenden Störungen: Welche Meldestrukturen existieren und wie ist das Meldewesen ausgeprägt?

[1] Hochschule Niederrhein University of Applied Sciences , Clavis - Kompetenzzentrum für Informationssicherheit, Webschulstraße 41-43, 41065 Mönchengladbach, Deutschland philipp.schuetz@hs-niederrhein.de

doi:10.18420/inf2019_61

5. Welche Standards, Normen und Rahmenwerke werden den Betreibern zur Härtung des informationstechnischen Sicherheitsniveaus empfohlen (z. B. ISO/IEC 27001, erweitert um sektorspezifische Ergänzungen)?
6. Welcher Verbindlichkeitscharakter entsteht durch entsprechende Betreiber-Sicherheitsnachweise (Vorlage eines Zertifikates, Auditierungen oder andere Nachweise)?

3 Erkenntnisse, Fazit und Ausblick

Von den eingangs formulierten sechs Fragestellungen können zum derzeitigen Stand die ersten vier Fragen hinreichend beantwortet werden. Mit Ausnahme von BE und den partiell zeitlichen Verzügen anderer Staaten, haben neun der zehn Staaten durch entsprechende Hoheitsakte die NIS-RL in nationales Recht überführt. Auch weisen die Ausprägungen der KRITIS-Strukturen mit den assoziierten Organisationen und Kontaktpunkten auf eine überwiegend strukturierte Detailierung hin. Die beiden abschließenden Fragestellungen nach konkretisierten Standards, Normen und Rahmenwerken, sowie den exakten Kontrollen und Nachweispflichten, können, nach jetzigem Stand und aus Sicht des Autors nur für DE und GB ausreichend beantwortet werden. Dies könnte, bezogen auf die acht Staaten, auf mindestens drei der folgenden Gründe zurückzuführen sein. Erstens: Die zuständigen Behörden erarbeiten noch entsprechende Branchenstandards und Vorgaben zu Rahmenwerken. Zweitens: Die Informationen sind nicht öffentlich zugänglich oder derzeit intransparent mit den betrachteten Dokumenten zur Umsetzung der NIS-RL verknüpft. Drittens: Die Vorgehensweise durch die Erhebung und Analyse der Primärquellen (Gesetze, Verordnungen und Strategien zur Cybersicherheit) reicht nicht aus, und so muss der Analyse-Radius möglicherweise um existierende Dokumente der jeweiligen Aufsichtsbehörden pro Sektor erweitert werden. Invariant von den im Beitrag geprüften Dimensionen zur Umsetzung der NIS-RL existieren in einigen Staaten Leitfäden, z. B. zum Umgang mit informationstechnischen Risiken. Daher werden dieser Arbeit weitere Untersuchungsaktivitäten folgen. Insbesondere mit dem Ziel, dass ein qualitativer Vergleich der Rahmenwerke für die Betreiber erfolgt und dadurch eine Synthese zu einem international anwendbaren Modell möglich wird. Dazu sollen zukünftig auch Länder außerhalb der Europäischen Union, wie z. B. die Vereinigten Staaten von Amerika und Japan in die Betrachtung einfließen.

Literaturverzeichnis

[SC19] Schütz, P.: Eine Übersicht und Vergleich zur NIS-Richtlinie innerhalb der EU mit dem Fokus auf die Betreiber von Kritischen Infrastrukturen, In: Tagungsband zum 16. Deutschen IT-Sicherheitskongress des Bundesamtes für Sicherheit in der Informationstechnik 2019, SecuMedia Verlag, Gau-Algersheim 2019, S.297-306

Track 5 – Sicherheit, Zuverlässigkeit, Korrektheit

Sicherheit, Zuverlässigkeit, Korrektheit

Juliane Krämer[1], Roland Meyer[2]

Der Track „Sicherheit, Zuverlässigkeit, Korrektheit“ verfolgt das Ziel, die Sicherheit und Verlässlichkeit von Systemen sowie von Anlagen und Betreibern zu erhöhen. Neben Vertraulichkeit, Integrität und Verfügbarkeit wird auch die Zuverlässigkeit im Sinne von Performanz, Fehlertoleranz und Korrektheit gegenüber einer Spezifikation betrachtet.

In dem Track werden verschiedene Systemklassen untersucht und sowohl theoretische als auch praktische Beiträge präsentiert. Die Beiträge erläutern unter anderem, wie in der Zukunft benötigte Kryptographie bereits heute praktisch eingesetzt werden kann, wie sich Fake-News von echten Nachrichten unterscheiden lassen und und wie neuartige Mobilitätskonzepte in Smart-Cities abzusichern sind.

Das Programmkomitee bestand aus:

- Peter Günther, Diebold Nixdorf
- Sven Jacobs, Universität des Saarlandes
- Anne Koziolek, KIT Karlsruhe
- Martin Leucker, Universität Lübeck
- Gerald Lüttgen, Universität Bamberg
- Kirsten Messer-Schmidt, excepture
- Isabel Münch, BSI
- Dirk Nowotka, Universität Kiel
- Jan Peleska, Universität Bremen
- Sebastian Schinzel, Fachhochschule Münster
- Marc Stöttinger, Continental Corporation
- Mario Trapp, Fraunhofer ESK München
- Heike Wehrheim, Universität Paderborn
- Bernhard C. Witt, it.sec GmbH & Co. KG

[1] Technische Universität Darmstadt, jkraemer@cdc.informatik.tu-darmstadt.de
[2] Technische Universität Braunschweig, roland.meyer@tu-bs.de

doi:10.18420/inf2019_62

Der Track erhielt neun Einreichungen, von denen vier akzeptiert wurden; die Annahmequote beträgt damit ca. 44 %. GI-Fellow Isabel Münch vom Bundesamt für Sicherheit in der Informationstechnik (BSI) konnte für einen eingeladenen Vortrag gewonnen werden. Zusätzlich präsentierten die Track-Chairs Juliane Krämer und Roland Meyer ihre Forschungsgebiete.

Full Papers

Post-Quantum Software Updates

A case study on Code Signing with Hash-based Signatures

Stefan-Lukas Gazdag[1] Markus Friedl[1] Daniel Loebenberger[2]

Abstract: Due to the progress in building quantum computers and the risk of attacks on cryptographic primitives based on quantum algorithms emerging, the development and analysis, but also the deployment of resistant schemes is an important research area. Hash-based signatures are a very promising candidate since they have been analyzed and improved for years. Nevertheless, there are some peculiarities that need consideration when using hash-based signatures in practice, for example the statefulness of some of the primitives. Fortunately, by now more and more experience is gained in real-world scenarios. In this paper we detail the troubles we encountered when using hash-based signatures in practice and study the most important use case for hash-based signatures: software or code signing.

Keywords: Post-Quantum Cryptography, Hash-based Signatures, XMSS, Software Update, ssh

1 Introduction

With the threat of quantum computing and its effect on today's cryptographic solutions arising, more and more companies and agencies are looking for alternatives to secure our current infrastructure. Well known schemes like RSA or the signature schemes DSA and its elliptic curve based equivalent ECDSA couldn't provide confidentiality, integrity nor authenticity when scalable quantum computer exist. For a survey on this topic see e. g. [BL17].

In the need for long-term security the transition to post-quantum schemes seems inevitable. Efforts have been made not only by the post-quantum cryptography community contributing and committing different schemes and providing security analyses but also by governmental and standardization agencies: In February 2017 the NSA [In17] released a recommendation to increase the security levels of currently used schemes and announced their plan to change to post-quantum schemes in the near future. NIST [Ch16; NI16] introduced a post-quantum crypto standardization project, the IETF has two RFCs on hash-based signatures (HBS) [Hu18; MCF19] as well as protocol-related efforts e. g. regarding IKEv2, and also ETSI [ET] is seeking for new post-quantum standards.

[1] genua GmbH, Kirchheim b. München, Germany, {stefan-lukas_gazdag,markus_friedl}@genua.eu
[2] Fraunhofer AISEC, Weiden i.d.Opf., Germany, daniel.loebenberger@aisec.fraunhofer.de

doi:10.18420/inf2019_63

The time needed for this transition is a major problem everyone has to face as the schedule is tight. Introducing new schemes is not only affiliated with building them and scrutinizing their security. They have to be standardized—which is currently a major effort—as well as integrated into software and tested in practical use cases.

As securing our whole infrastructure may take longer than the time it takes to build a large enough quantum computer (for a treatise of this issue we refer to [Mo15]), users long for introducing post-quantum cryptography to the first use cases where it is applicable. One of the use cases we can safeguard today is code signing and software updates. These are critical for a secure migration to a fully post-quantum secure system. If quantum attacks occur earlier than expected and systems are not yet equipped with a quantum-resistant update mechanism, there is no easy way to secure the systems afterwards. In the worst case a manufacturer would have to deliver updates with a secure carrier.

But even in the case that quantum computers can never be built due to foundational problems not being known or understood today, work motivated by the transition to post-quantum schemes will have great benefit for our understanding of applied cryptography. Furthermore, a diverse range of cryptographic schemes based on different kinds of problems would enhance our knowledge, since a versatile pool of well-analyzed, practically relevant and deployed schemes would supply enough alternatives to switch to another scheme if trouble arises for one of the currently used ones. For this a lot of questions are yet to be answered, e. g. regarding modular design of protocols, hybrid solutions for using cryptography and testing protocols and implementations while being confronted with so many different combinations of cryptographic schemes.

Thus, it makes sense today to experimentally employ selected post-quantum primitives in practice. We are in good company here: Google announced in mid 2016 that they would run first real world experiments with post-quantum cryptography. In the announcement they note [In16]:

> "Our aims with this experiment are to highlight an area of research that Google believes to be important and to gain real-world experience with the larger data structures that post-quantum algorithms will likely require."

They have pursued their efforts since and others have followed.

Motivation. But which post-quantum primitives come into consideration for such experiments? As stated above, we need a scheme which is well analyzed and standardized. What we mostly do have are experimental schemes which are in a process of standardization, notably the candidates [NI17; NI19] of the NIST call [NI16] for post-quantum cryptographic algorithms and the hash-based signatures from the informational RFCs, namely XMSS [Hu18], Leighton-Micali Signatures, and HSS [MCF19].

In this article we discuss a specific use case which is using hash-based signatures in the context of software updates. The cryptographic foundations of these signatures were invented quite early by Merkle in 1979 [Me79] and advanced in works like [BDH11; BDS08; Bu07; HRB13]. Unfortunately, the schemes suffer some peculiarities (the size of cryptographic keys, statefulness of the private key, etc.) which ruled them out for classical applications for some time. On the other hand they now look promising in the post-quantum context. Most other post-quantum schemes are comparatively recent academic developments which clearly have not went through thorough analyses by many people over several decades, though a lot of work is invested to enhance our knowledge. Still, the trust in those schemes is not fully established yet and it seems too early to apply them broadly in practice, though in some cases we already have to do so (e. g. by using hybrid solutions) to gain experience and trust.

Additionally, for hash-based signatures there are also first results on practical implementation issues [HBB12; KMN14]. For LMS there exists RFC 8554 [MCF19] and XMSS was released as RFC 8391 [Hu18]. The work of several authors showed how crucial a secure environment is to the handling of the private key [CMP18; FG18; Ge18; Ka18], yet we know how to deal with this in the case examined in this article. What is still missing, is work on the correct use of these algorithms in real-world settings. To do so we want to exhaustively describe the most suitable use case for hash-based signatures and the surrounding settings in which to use these. We also evaluate the features and stress the limitations of this kind of schemes.

2 Peculiarities of hash-based signatures

2.1 Hash-based signatures

We first give a brief introduction to hash-based signatures partly following [Mc16]. Unlike most other signature schemes, HBS require only a secure cryptographic hash function and no other hardness assumption (about a number-theoretic problem) and are not known to be vulnerable to Shor's algorithm. *Secure* here refers to either collision resistance or mere second-preimage resistance, depending on the specific construction.

One-time signature schemes. Hash-based signatures use one-time signature (OTS) schemes as a fundamental building block. Common examples are the seminal one by Lamport [La79], the Winternitz scheme [DSS05], and its recent variant W-OTS$^+$ [Hü13]. For descriptions of current one-time signatures we refer to the respective section of [Hu18; MCF19]. For one-time signatures, the private key is usually generated randomly and the public key is a function of the private key, involving the underlying hash function. Advanced one-time signature schemes feature a parameter enabling a time/memory trade-off, e. g. the Winternitz parameter. One-time schemes are inadequate on their own in practice, since each private key can only be used to securely sign a single message.

Hash-based or N-time signature schemes. Merkle introduced a way to manage a confined number of OTS key pairs in 1979 [Me79] by using a binary tree to administer the one-time keys. A tree with height H may hold up to $N = 2^H$ key pairs and sign this number of messages. While OTSs use hash functions just as well, it is this construct that is referred to as *hash-based signatures*. It consists of the classical algorithms of key generation, signing and verification. Several improvements have been made for hash-based signatures, e. g.[BDH11; BDS08; Bu07; LM95]. For each signature generation a different OTS key pair is used. An integer counter has to keep track of the advancement of the key. A simple way to reduce the size of an N-time key is to define it to be a short string and then use a cryptographically secure pseudorandom function to generate the actual keys of the underlying one-time scheme.

Hierarchical Signatures. A *hierarchical signature scheme* is an N-time signature scheme that uses other hash-based signatures in its construction. In literature it is mostly referred to as *multi-tree* or *hyper-tree* variant of classical hash-based signatures. A scheme of this kind uses layers of trees meaning the root of a single tree is signed by an OTS key of a different tree. That way trees of large height may be build with relatively small cost in performance and memory demand. Concrete examples of hierarchical hash-based signatures include XMSS^{MT} [HRB13], a scheme by Leighton and Micali [LM95] and SPHINCS [Be15]. XMSS^{MT} and SPHINCS define a parameter d as the number of layers in the hierarchical structure. Additionally, the LMS [MCF19] specification describes a hierarchical hash-based signature variant based on Leighton and Micali's scheme, called HSS.

2.2 Statefulness

Hash-based signatures introduce some rare properties. The most important one is the *statefulness* of the private key of most hash-based signature schemes: A secret key has to be updated with the generation of each signature, since the underlying one-time signature scheme must only use its private keys exactly once to preserve its security features. In the most basic case it is the index of the next OTS key pair within the Merkle tree which has to be updated. In advanced implementations using pseudo-random generation of the OTS key pairs and an improved tree traversal algorithm also the seed for the pseudo-random number generator and nodes within the tree have to be updated in addition.

Another issue is the limited number of OTS key pairs. A private key exhausts due to the limited number of signing keys. Therefore a decision about the required number of signing keys must be made before key generation. And for the case that the HBS key has run out of OTS keys, some form of warning and key exchange system has to be installed.

The private key therefore needs special consideration and attention. No matter if the key is stored unguarded e. g. on a hard disk (which we definitely do not recommend) or if some form of key management was established the private key has to be seen as a critical resource.

An exclusive access to the private key must be enforced. Multiple processes or instances must not access the private key at the same time. Indeed, these issues regarding statefulness were resolved or easened, e. g. with a reservation approach [Mc16].

Yet some issues can't be addressed in common ways. Any copy of the secret key may be problematic. As [Mc16] show it is quite some effort already to make sure no unwanted copies are left within a modern computer system. The wish for having a backup of the secret key is tough as this would mean one would have to update the backup key as well. Some approaches as for state management might be applied to a backup as well, but still lead to complex scenarios. We rather recommend to provide several stateful keys and to deploy their public keys altogether. Then all but one of the secret keys may remain unused and stored in a secure place. If there's a problem with the currently used key one may switch to the "backup" key(s). Traditional backup mechanism may only be applied to stateless schemes.

Stateless schemes do exist, the first being SPHINCS [Be15], followed by the two submissions to the NIST standardization process [AE17; Be17]. Instead of one-time signature schemes so-called few-time signature schemes are used. With a single key pair multiple signatures may be generated with each signature generation reducing the security. That way the index of the key pair used can be chosen at random. Signatures of such kind have bigger sizes, while the need for handling a state is exchanged with a probability calculation.

2.3 API incompatibility

The need for an update of the private keys implies that HBS typically don't match common interfaces of signing operations of state of the art cryptographic tools and libraries. Even more, stateful HBS don't fit the definition of cryptographic signature schemes in general, since those traditionally only consist of `keygen` for key generation, `sign` for signing a message with a private key and `verify` for verification of a signature given a public key. Now another operation is needed: the operation `update` which evolves the key. One solution is to implement the function `update` implicitly as part of the signing operation `sign`.

McGrew et al. [Mc16] propose a state reservation method which grants access to a specific state of the key. Unfortunately, every way of implementing a secure form of HBS key management will always have some overhead due to writing operations or special key update tactics.

2.4 Storage and speed requirements

For different use cases various key sizes and different forms of key management might be suitable. Even a pool of HBS keys would be conceivable. This in turn has a vast effect on the amount of memory needed for the key material, the signature itself and many temporary

data structures which are employed during the signing/verification/update procedures. Also, the time required for all operations might vary considerably.

In the case of hash-based signatures, there is a natural trade-off between storage size and life-time of the secret key. A small key will have a short life-time, while a large key can be used more often. Also, the time needed for different cryptographic operations depends on the size of the key: Typically, the time increases when the keysize grows.

3 Software Update Authentication

We now explore one of the most suitable use cases for HBS: the authentication of software updates or firmware. When delivering IT systems or software, manufacturers usually provide means of authenticity. This could be a sealed package handed to the customer or a software being installed only if it was signed by a key an operating system already knows. Once the user trusts the system, the different applications take care of further updates.

A non-specified attacker (e. g. an intelligence agency) may be able to forge signatures for a malicious product update and include a backdoor. To prevent attacks, software update mechanisms as well as the updates themselves must be secured. This is especially important when systems need to run over a long period of time. Many servers and machines are working for years, sometimes even decades. For instance, satellites may be updated via a wireless connection. However, while improving parts of a system via updates may be straightforward, modifying currently used update mechanisms on older machines or systems remotely, as satellites sometimes remain active for a long time, can be tricky.

Therefore, a mechanism and parameter settings have to be used that provide security for at least the expected period of service. Of course advances in cryptanalysis and software bugs may require a change of the currently used system. Therefore it is always advisable to design the update mechanism as modular as possible and to allow easy exchange or update of keys or the mechanism itself.

As noted, hash-based signatures are particularly well suited for software updates: On one hand, updates are often released in an interval of at least days (e. g. to apply security fixes), weeks (patches for newly released computer games) or even months (major product updates). Thus, comparatively few signatures need to be generated. Even a small key with a tree of height 10 lasts for 85 years assuming monthly updates. A key using a tree of height 20 would last 29 years, even with a hundred signatures per day (e. g. including test builds). Another way to deal with this is to e. g. use one key pair per major release if this suits the update policy.

On the other hand, in many update settings, neither key generation nor signature generation or verification are too time-critical. Only the last may be limited e. g. on resource-restricted devices, but a tailored solution with minimal verifier (only providing mandatory or necessary parameter sets) with minimal code base should offer adequate performance and compatibility.

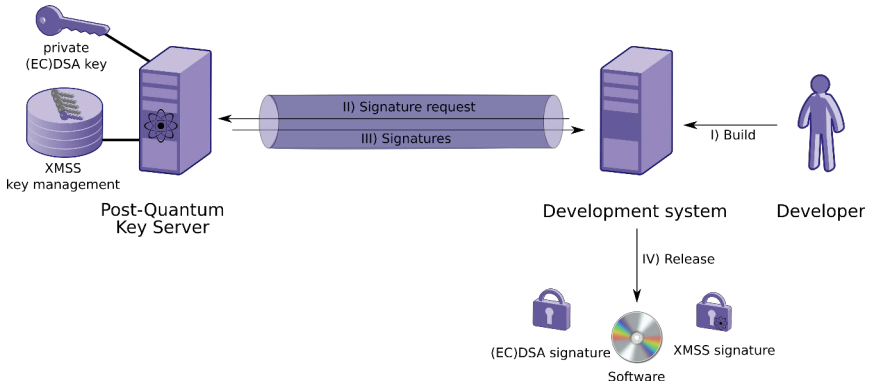


Fig. 1: A simple but secure setup for software development. A key server connects to selected development machines and handles requests for signing code.

A key pair may be generated in advance, so the process could take hours without causing trouble. Usually it does not matter if the signature generation or verification takes a few hundred (milli-)seconds. There might be some restrictions in size, depending on the specific update mechanism, the used data structures and possibly memory restrictions on devices with restricted resources. Luckily the public key of most HBS is quite small. Only the signature size might be problematic in few cases. Also, secure updates are universally relevant, since they may serve as a basis for realizing secure cryptographic algorithm-agility.

3.1 Software Update Environments

We now examine the process of releasing a signature for a software update. Numerous ways of implementing a software development environment and update mechanisms exist, one particular way is depicted in Figure 1. A typical infrastructure consists of a server (or a distributed net of servers) holding the current version of the code, a dedicated build server (which is sometimes in union with the code server) and a key server. Especially large companies use more advanced systems and tools to control and manage huge amounts of code and software, where lots of different products or services are managed, instead of a single product (version) as described here. Ideally, changes to the code base involve at least three persons: A software developer working on the actual change (two considering pair programming), at least one reviewer checking the code for bugs, conceptual errors as well as backdoors or other malicious changes and an integrator who includes the change (ideally after another review) into the current code base. Finally, a build server takes the cleared source code and builds a final package to be released or for test purposes. Such a software package is to be cryptographically secured e. g. by signing the content of the package.

To this end, the build server needs a signing key which resides on a distinct key server. The key server may hold diverse keys, but each key should be only used for one single product or

a specific purpose. Ideally, each key is stored on a dedicated smartcard or at least provided by a different security anchor like a hardware security module. The communication between those systems is to be secured, even if it appears to be a separated network. This can be done easily via TLS, SSH or other protocols securing the connection. To build a code package, the build server takes the latest code base, builds the software, asks the key server to sign e. g. a message digest respectively a hash value of the package using a particular signing key and hands out the package to be released as well as the signature. The targeted system is now able to verify the integrity and issuer of the package before installing the update.

A pleasant feature of update mechanisms is the fact that in many cases, it is possible to establish a hybrid solution which may help increasing confidence in the use of HBS. An update mechanism could use ECDSA [In13] or EDDSA [JL17] to verify the software and double-check by using an XMSS key and signature. This allows us to add signatures that are believed to be quantum-safe while still relying on well-established cryptographic primitives such as (EC)DSA.

In some cases companies sign the same update with several keys to offer compatibility for upgrading from older machines, though this may contain steps meaning you may have to install a version in-between first before going to the targeted version. In such an environment HBS could be simply another rollout of an advanced key. No matter whether a hybrid or a successive solution is used, in a pre-quantum era, it is possible to only trust the classical key at first and observe the behavior of the post-quantum part of the update mechanism. In the unlikely event of a deviation in the verification process, a skeptical user (and with the assistance and allowance of the user the manufacturer) could check for the cause. This way HBS can be deployed, show their qualification as successor technology and be switched to on the way before attacks by quantum computers emerge.

3.2 Our Implementation

We now discuss the integration of HBS in a currently used update mechanism and its implications, bearing stepping-stones that have been shown before. The main drawback of HBS is to be found in the statefulness of some schemes. Fortunately, for update mechanisms, this disadvantage dissipates. Even relatively small tree sizes (e. g. a tree of height 20) suffice to sign daily test builds and official software or patch releases. Typical software updates in our case have sizes of a few to many megabytes. Often it does not matter if a signature requires a few kilobytes. Speed is not that relevant either, as a user installing the update will not notice a few more hundred milliseconds. As a consequence, issues typical for HBS have less impact here. The statefulness mainly affects the key server which has to offer a key management instance that controls the access to the actual key, like the reservation approach proposed by McGrew et al. [Mc16]. While in a regular case one would directly call the signing function, for HBS the reservation function is called first. It then returns the current state of the key.

Though some quirks of HBS seem not to be relevant, some aspects still need consideration. One does have to think about dependencies by (automatic) test setups. While testing the signature generation and verification may not be troublesome considering the runtime, the generation of a key might be. This can take several minutes even for fairly small parameter sets and result in the delay of other tests.

Another issue is providing a secure environment, which must not be neglected for classical schemes as well. It is tricky to actually enforce such an environment on off-the-shelf hardware. One option is to employ smartcards, which are an important security anchor for providing cryptographic services while protecting the key material. In most cases, where a central secure key server is not suitable, a smartcard would be the only really secure way to use HBS. But the nature of a smartcard is not well-disposed for the use of HBS. In conventional cases, a secret key on a smartcard only has to be read after it was written to the memory. With HBS, some amount of data has to be rewritten, which might cause trouble. The first constraint is that a smartcard with re-writable memory must be used, but the memory wears out over time with each writing operation. As a result, the rewriting of data has to be managed to avoid that some sections of the memory are used too often and may get broken. Hülsing et al. [HBB12] have shown how smartcards can handle HBS schemes like XMSS^{MT}. If a smartcard is not applicable, yet a secure environment has to be used. First Hardware Security Modules (HSMs) are available with experimental support for quantum-resistant schemes. At least a hardened server with corresponding restrictions has to be used, so an attacker has no influence especially on the key generation.

We selected XMSS as described RFC [Hu18] for our implementation, since it suits our security requirements well. To do so, a parameter set should be picked from the options provided by the IETF specification according to personal needs: A total tree height of 20 was sufficient for most applications in our setup, independently of whether XMSS or XMSS^{MT} is used, though bigger tree height could be used easily. Relative to other official parameter sets, this is a large height for XMSS but a small one for XMSS^{MT}. We recommend XMSS^{MT} for performance reasons, but XMSS can also be used and has the advantage of a simpler implementation. In practice, we recommend

- `XMSS-SHA2_20_256` or `XMSS-SHAKE_20_256` for XMSS and
- `XMSSMT-SHA2_20/2_256` or `XMSSMT-SHAKE_20/2_256` for XMSS^{MT} for a more efficient implementation.

For a higher security level and being well prepared for a post-quantum era one should opt for

- `XMSS-SHA2_20_512` or `XMSS-SHAKE_20_512` for XMSS
- `XMSSMT-SHA2_20/2_512` or `XMSSMT-SHAKE_20/2_512` for XMSS^{MT}.

Note that some of these are optional parameter sets in RFC 8391.

To secure the connection between the systems involved, we decided to access the servers via OpenSSH and implemented experimental XMSS support[1]. Also other post-quantum suites are tested to establish a quantum-resistant connection.

For the purpose of signing data, several open-source applications exist, e. g. in the OpenBSD context `signify` or its predecessor `gzsig`. We opted for the latter due to its greater flexibility and simplicity. These tools are easily adaptable by introducing the use of XMSS via X.509 certificates or SSH keys and including the XMSS functionality by either using an adapted cryptographic library like LibreSSL[2] or forks with a focus on quantum-resistant algorithms like the Open Quantum Safe Project[3] or at least by some verified standalone solution. A standalone solution would make it possible to offer a minimal verifier for the updated systems, but also means that it is an individual solution for that product. In our case we use OpenSSH to offer all cryptographic functionality.

On start-up a secure key server connects to several development machines. Access to the key server is restricted by digital as well as physical means. As soon as a signature is needed for a new patch, the key server is asked to sign the code package by the signing tool. As usual instead of the whole code package a hash value is signed. The key server contains the reservation-based key management and controls the access to the XMSS keys and other private keys. A signing agent awaits the requests and though only one signature per signature scheme used is needed, the signing agent uses the reservation function to reserve an interval of keys. In this case the software solution or cryptographic library used to supply XMSS functionality on the build server or in software itself (verifier) does not need to provide the state management instance, as an instance on the key server takes care of this. In addition, digests (e.g. SHA2-512 or SHA3-512) of the package or signatures may be provided via other channels than the package itself. One alternative is to provide package digest values to users via a web interface. A software system to be updated holds a public key that can be used to verify the new software or update package (e.g. in the X.509, SSH or PKCS8/12 format).

As described before, the transition to post-quantum cryptography is best done by implementing hybrid solutions. We highly recommend to do so in the case of adapting or extending update mechanisms. In many cases, it should be straightforward to verify multiple signatures of different signature schemes and check for the correct verification of all tested signatures. Nevertheless some settings might include resource-restricted environments or deployed software that can't be updated easily. At least it is possible with most HBS to offer a minimal verifier while also having a quite small public key. This easens for example the requirements of the IoT or embedded devices. We use ECDSA as a pre-quantum solution which might resist first attacks based on early quantum computers as standard case to trust, while also verifying an XMSS signature. During an initial transition phase, both ECDSA and XMSS signatures could be verified in parallel with only the result of the ECDSA

[1] https://www.openssh.com/txt/release-7.7; git repository https://github.com/openssh/openssh-portable
[2] https://www.libressl.org/
[3] https://openquantumsafe.org/

Device	Type
Processor	AMD Geode LX800
Speed	500MHz
Memory	256MB SDRAM
Operating System	OpenBSD 6.1

Object/Operation	Requirements
Secret key	4363 Byte
Public key	190 Byte
Signature	2820 Byte
Key generation	25 hours
Signing	0.91s
Verification	0.75s

Tab. 1: Some characteristics of our key-server (left) as well as the corresponding benchmarking results (right). Shown are values for the variant `XMSS_SHA2_20_256` following RFC 8391. The keys were stored in the ssh key format. Only approximate timing results are given, since the verification time is client-dependent. Note that the results show sizes including the encoding overhead and times include any software overhead.

verification actually being taken into account. Nevertheless we opted for trusting both signatures, classical as well as quantum-resistant, so no update can be installed if a single one of them fails verification. Introducing this approach to an update mechanism can mean that no downgrade or fallback to a former software release is possible unless the update mechanism would not check the new signatures if an old version is to be installed. This is a serious attack vector as an adversary might forge a software patch masqueraded as an old version. Thus we recommend to only accept with both signatures verifying positively and in our case we have not experienced or heard about any problems even on live systems in the field. We expect that even with sticking to an optional transition face confidence in HBS will increase gradually and remove any impact on actual verification in the unlikely event of an XMSS verification failure.

Also, a hybrid approach constitutes a secure solution to comply to conflicting governmental requirements. Imagine various agencies asking for different cryptographic signature schemes to be used in specific settings. Thanks to the hybrid approach several schemes may be used targeting diverse specifications while each provider of requirements can be sure that each update is checked by means of own liking.

If small trees shall be used or backup keys are wished for other reasons a key management may provide several keys. Each major release may be shipped with several keys, so there's one active signing and verification key while there's e. g. two backup keys available. If no unforeseen events enforce the exchange of more than one key, the active key may be replaced with the next major update. This also works for a backup key server. If a problem with the running key server occurs one might switch to a backup system or restore the actual key server which can use the unused, securely (and depending on the solution independently) stored extra keys.

4 Conclusion

We presented a working implementation of a post-quantum secure software update mechanism based on hash-based signatures. The code-signing mechanism is currently run in a hybrid fashion employing ECDSA and XMSS signatures at the same time.

This allows us to experimentally work with novel cryptographic algorithms resistant to quantum computers in practice without compromising the well-established security of the update mechanism. Our implementation was carried out on an OpenBSD-based kernel and a version of OpenSSH providing basic XMSS support for the purpose of post-quantum secure software updates.

Summarizing, we were able to realize post-quantum secure crypto-agility in practice, offering real-world software signatures that cannot be broken by a scalable quantum computer—as far as we know. Such a protection for software updates can and should be used today in components for high-security areas with long expected deployment times.

Acknowledgments

The work underlying this article was mainly done when the third author worked with the first two authors at genua GmbH, Kirchheim, Germany. We thank Hans-Jörg Hoexer for collecting the benchmark-data given in Table 1 and Tobias Heider for his review.

References

[AE17] Aumasson, J.-P.; Endignoux, G.: Gravity-SPHINCS, Submission to the NIST standardization process, 2017.

[BDH11] Buchmann, J.; Dahmen, E.; Hülsing, A.: XMSS — A Practical Forward Secure Signature Scheme Based on Minimal Security Assumptions. In: PQCrypto. Vol. 7071. LNCS, Springer, pp. 117–129, 2011, ISBN: 978-3-642-25404-8.

[BDS08] Buchmann, J.; Dahmen, E.; Schneider, M.: Merkle Tree Traversal Revisited. In: PQCrypto. Vol. 5299. LNCS, Springer, pp. 63–78, 2008, ISBN: 978-3-540-88402-6.

[Be15] Bernstein, D. J.; Hopwood, D.; Hülsing, A.; Lange, T.; Niederhagen, R.; Papachristodoulou, L.; Schneider, M.; Schwabe, P.; Wilcox-O'Hearn, Z.: SPHINCS: Practical Stateless Hash-Based Signatures. In: EUROCRYPT 2015. Vol. 9056. LNCS, Springer, pp. 368–397, 2015, ISBN: 978-3-662-46799-2.

[Be17] Bernstein, D. J.; Dobraunig, C.; Eichlseder, M.; Fluhrer, S.; Gazdag, S.-L.; Hülsing, A.; Kampanakis, P.; Kölbl, S.; Lange, T.; Lauridsen, M. M.; Mendel, F.; Niederhagen, R.; Rechberger, C.; Schwabe, P.: SPHINCS+, Submission to the NIST standardization process, 2017.

[BL17] Bernstein, D. J.; Lange, T.: Post-quantum cryptography. Nature 549/, pp. 188–194, 2017.

[Bu07] Buchmann, J.; Dahmen, E.; Klintsevich, E.; Okeya, K.; Vuillaume, C.: Merkle Signatures with Virtually Unlimited Signature Capacity. In: ACNS. Vol. 4521. LNCS, Springer, pp. 31–45, 2007, ISBN: 978-3-540-72737-8.

[Ch16] Chen, L.; Jordan, S.; Liu, Y.-K.; Moody, D.; Peralta, R.; Perlner, R.; Smith-Tone, D.: Report on Post-Quantum Cryptography (NISTIR 8105), http://nvlpubs.nist.gov/nistpubs/ir/2016/NIST.IR.8105.pdf, Accessed 2017-02-20., 2016.

[CMP18] Castelnovi, L.; Martinelli, A.; Prest, T.: Grafting Trees: a Fault Attack against the SPHINCS framework. In: International Conference on Post-Quantum Cryptography. Springer, pp. 165–184, 2018.

[DSS05] Dods, C.; Smart, N. P.; Stam, M.: Hash Based Digital Signature Schemes. In (Smart, N. P., ed.): Cryptography and Coding. Vol. 3796. LNCS, Springer, pp. 96–115, 2005, ISBN: 3-540-30276-X.

[ET] ETSI: Quantum-Safe Cryptography, http://www.etsi.org/technologies-clusters/technologies/quantum-safe-cryptography, Last access on 2017-04-18.

[FG18] Fan, J.; Gierlichs, B., eds.: Constructive Side-Channel Analysis and Secure Design - 9th International Workshop, COSADE 2018, Singapore, April 23-24, 2018, Proceedings, vol. 10815, Lecture Notes in Computer Science, Springer, 2018, ISBN: 978-3-319-89640-3, URL: https://doi.org/10.1007/978-3-319-89641-0.

[Ge18] Genêt, A.; Kannwischer, M. J.; Pelletier, H.; McLauchlan, A.: Practical Fault Injection Attacks on SPHINCS, Cryptology ePrint Archive, Report 2018/674, https://eprint.iacr.org/2018/674, 2018.

[HBB12] Hülsing, A.; Busold, C.; Buchmann, J. A.: Forward Secure Signatures on Smart Cards. In: SAC. Vol. 7707. LNCS, Springer, pp. 66–80, 2012, ISBN: 978-3-642-35998-9.

[HRB13] Hülsing, A.; Rausch, L.; Buchmann, J.: Optimal Parameters for XMSS^{MT}. In: Security Engineering and Intelligence Informatics — CD-ARES 2013 Workshops: MoCySEn and SeCIHD. Vol. 8128. LNCS, Springer, pp. 194–208, 2013, ISBN: 978-3-642-40587-7.

[Hü13] Hülsing, A.: W-OTS$^+$ — Shorter Signatures for Hash-Based Signature Schemes. In: AFRICACRYPT. Vol. 7918. LNCS, Springer, pp. 173–188, 2013, ISBN: 978-3-642-38552-0.

[Hu18] Huelsing, A.; Butin, D.; Gazdag, S.-L.; Rijneveld, J.; Mohaisen, A.: XMSS: eXtended Merkle Signature Scheme, RFC 8391, IRTF, May 2018, 74 pp., URL: https://rfc-editor.org/rfc/rfc8391.txt.

[In13] Information Technology Laboratory: Digital Signature Standard (DSS), en, tech. rep. NIST FIPS 186-4, National Institute of Standards and Technology, July 2013, URL: https://nvlpubs.nist.gov/nistpubs/FIPS/NIST.FIPS.186-4.pdf, visited on: 04/12/2019.

[In16] Inc., G.: Experimenting with Post-Quantum Cryptography, https://security.googleblog.com/2016/07/experimenting-with-post-quantum.html, July 2016.

[In17] Information Assurance Directorate at the National Security Agency: Commercial National Security Algorithm Suite, https://www.iad.gov/iad/programs/iad-initiatives/cnsa-suite.cfm, Accessed 2017-02-20., 2017.

[JL17] Josefsson, S.; Liusvaara, I.: Edwards-Curve Digital Signature Algorithm (EdDSA), RFC 8032 (Informational), RFC, Fremont, CA, USA: RFC Editor, Jan. 2017, URL: https://www.rfc-editor.org/rfc/rfc8032.txt.

[Ka18] Kannwischer, M. J.; Genêt, A.; Butin, D.; Krämer, J.; Buchmann, J.: Differential Power Analysis of XMSS and SPHINCS. In (Fan, J.; Gierlichs, B., eds.): Constructive Side-Channel Analysis and Secure Design - 9th International Workshop, COSADE 2018, Singapore, April 23-24, 2018, Proceedings. Vol. 10815. Lecture Notes in Computer Science, Springer, pp. 168–188, 2018, ISBN: 978-3-319-89640-3, URL: https://doi.org/10.1007/978-3-319-89641-0_10.

[KMN14] Knecht, M.; Meier, W.; Nicola, C. U.: A space- and time-efficient Implementation of the Merkle Tree Traversal Algorithm. CoRR abs/1409.4081/, 2014.

[La79] Lamport, L.: Constructing Digital Signatures from a One Way Function, tech. rep., https://www.microsoft.com/en-us/research/publication/constructing-digital-signatures-one-way-function/ Accessed 2017-02-20., SRI International Computer Science Laboratory, 1979.

[LM95] Leighton, T.; Micali, S.: Large provably fast and secure digital signature schemes from secure hash functions, U.S. Patent 5,432,852, 1995.

[Mc16] McGrew, D. A.; Kampanakis, P.; Fluhrer, S. R.; Gazdag, S.-L.; Butin, D.; Buchmann, J. A.: State Management for Hash-Based Signatures. In: SSR. Vol. 10074. LNCS, Springer, pp. 244–260, 2016, ISBN: 978-3-319-49099-1.

[MCF19] McGrew, D.; Curcio, M.; Fluhrer, S.: Leighton-Micali Hash-Based Signatures, RFC 8554, IRTF, Apr. 2019, 61 pp., URL: https://rfc-editor.org/rfc/rfc8554.txt.

[Me79] Merkle, R. C.: Secrecy, Authentication and Public Key Systems, PhD thesis, Dept. of Electrical Engineering, Stanford University, 1979.

[Mo15] Mosca, M.: Cybersecurity in an era with quantum computers: will we be ready?, Cryptology ePrint Archive, Report 2015/1075, https://eprint.iacr.org/2015/1075, 2015.

[NI16] NIST: Federal Register Vol. 81, No. 244, Dec. 2016.

[NI17] NIST: Post-Quantum Cryptography: Round 1 Submissions, https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Round-1-Submissions, Last access on 2018-03-23, Mar. 2017.

[NI19] NIST: Post-Quantum Cryptography: Round 2 Submissions, https://csrc.nist.gov/Projects/Post-Quantum-Cryptography/Round-2-Submissions, Last access on 2019-04-11, Jan. 2019.

Quotable Signatures using Merkle Trees

Michael Kreutzer[1] Ruben Niederhagen[1] Kris Shrishak[2] Hervais Simo Fhom[1]

Abstract: Fake news have been around since time immemorial. But the widespread reach and the rate of propagation through social media websites makes the issue of fake news a grave concern. We propose to address the issue of fake news through the use of quotable signatures using Merkle trees to verify news shared on social media websites.

Keywords: Fake news, Merkle tree, social media.

1 Introduction

Recently, there has been a growing concern about the spread of fake news. Though the phenomenon of fake news is not new, the use of social media websites allows for the fake news to reach a wider audience in a shorter span of time, thus necessitating to suppress them at an early stage. Though there are multiple methods used to spread fake news, in this work, we focus on the following scenario: A journalist at a news organization researches and writes an article and publishes it on the organization's website. Reading this article, a user of a social media website wishes to share it with his/her followers by quoting a part of the text from the news article. Since the origin of the text is from an established organization and since the topic has been well researched by the author, the user wants to include a digital signature of the author or the organization in order to proof the origin of the text. This allows the community to detect well researched contents more easily and to distinct it from fake news. For quoting parts of a text while maintaining its signature, we propose "quotable signatures" using Merkle trees, which is simpler than previously proposed solutions based on machine learning [RSL17; SFR17].

2 Fake News Detection

Our goal is to have a single signature that is valid for the entire text but also for partial quotes of the text. In order to achieve this goal, we propose to use a Merkle tree [Me79]: The input text is split into single words and punctuation marks. Hashes of these tokens are

[1] Fraunhofer Institute for Secure Information Technology, Darmstadt, Germany.
ruben.niederhagen@sit.fraunhofer.de
[2] TU Darmstadt, Darmstadt, Germany.

doi:10.18420/inf2019_64

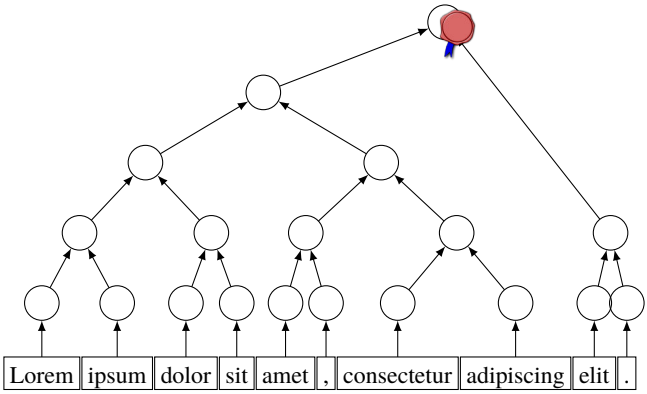


Abb. 1: Merkle signature tree for a short example.

the leaves of the Merkle tree. The Merkle tree over these leaf nodes is a binary hash-tree. The nodes on each level are pairwise-hashes of nodes on the previous level. For computing a quotable signature, the author computes the Merkle tree up to the root node and signs this root with his/her private key. Figure 1 shows an example of a Merkle signature tree for a short example text. The signature of the text can be easily checked by a verifier by re-computing the root of the Merkle tree and by verifying the signature on the root.

When quoting an article, naturally only a (small) part of the original text is provided. Therefore, not all the leaf nodes of the Merkle tree are available and the verifier requires additional information for computing the root node and thus for verifying the signature. However, instead of providing the entire text, it is sufficient to provide the *verification path* in the Merkle tree, i.e., those nodes in the tree that are missing on the path from the quoted text nodes to the root node. Figure 2 shows an example of the verification path if only a few root nodes are provided in the quote. In this example, the verification path consists of the dark-gray nodes. Given the quoted text and these verification nodes, the verifier is able to re-compute the root node and to verify the signature.

The leaf nodes are computed as hashes of the text tokens. The inner nodes are computed as hashes of the concatenated values of their child nodes: The value of the right child node is appended to the value of the left child node. This guarantees that the order of the words in the quote can not be changed by an attacker. Due to the structure of the Merkle tree, the verifier knows which leaf nodes are missing in the quoted text and therefore has information about where the quote is skipping parts of the original text and therefore where context might be missing.

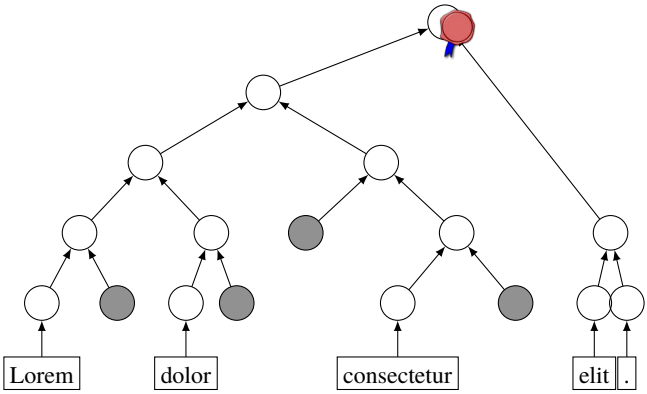


Abb. 2: The gray nodes give the verification path for a set of leaf nodes.

2.1 Cost

Usage of quotable signatures imposes some cost in terms of computation and data size (storage and transmission). These costs occur for signers (authors) that provide quotable signatures for their text, for *quoters* who quote some signed text, and for verifiers (readers) who want to verify quotable signatures on original texts or quotes.

Signers. The computational overhead for signers consists of the computation of the Merkle tree and the generation of the signature on the root node. For a text of n tokens, the height of the tree is $h = \lceil \log(n) \rceil$. There are $2^h - 1$ inner nodes that require concatenated hashes from their child nodes and n leaf nodes that are hashes of the text tokens. Thus, in total, the signer has to compute $2^h + n - 1$ calls to the hash function plus the computation of the signature on the root node.

The increment of the data size for a complete text with quotable signature is moderate. The Merkle tree does not need to be published; only the signature on the root node is needed in addition to the original text. Additional information is required depending on the certification scheme.

Quoters. The cost for the quoters depends on the number of words quoted and the number of sections quoted. For estimating the worst case cost for the quoter, consider the following three cases:

- *One single word* is quoted:
  The entire verification path from the leaf node up to the top needs to be provided.

Therefore, the verifier has to recompute the entire Merkle tree, i.e., $2^h + n - 1$ hashes. The entire verification path of h nodes needs to be provided; the data size grows by h hashes.

- *One single, continuous part* of the text is quoted:
  The exact cost depends on the number of tokens that are quoted. The worst case is a quote of two words in the middle of the Merkle tree, requiring independent verification paths almost up to the root node. In this case, the quoter needs to compute $2^h - 1 - 2\lceil \log_2(n/2) \rceil$ hashes. The number of nodes in the verification path is $2\lceil \log_2(n/2) \rceil$.
- *Several parts* are quoted:
  The number of nodes that need to be re-computed in the Merkle tree shrinks the more tokens are quoted. However, the worst case for the data size of the verification path is when every second word is quoted. This requires $n/2$ nodes in the verification path. (In this case, it is more efficient to provide the original tokens instead of their hashes.)

Therefore, the worst case with regards to computational cost is the re-computation of almost the entire Merkle, i.e., $2^h - 1 - \lceil \log_2(n) \rceil$ hashes. The worst case for the data size is to quote every second word, resulting in $n/2$ hashes.

Verifiers. In the worst case, the verifier needs to re-compute the entire Merkle-tree with n leaf nodes and $2^h - 1$ inner nodes This results in a cost of $2^h + n - 1$ calls to the hash function. The more nodes are provided in the verification path, the less hashes need to be re-computed by the verifier. In any case, the verifier needs to verify the signature on the root node and the certificate of the signer.

2.2 Statistics and Efficiency

In 1951, Shannon claimed that the average word length in English texts is 4.5 characters [Sh51]. Recent studies show that in modern texts the average word length is slightly larger[3]. In the following, we use 4 characters as the lower bound for a worst-case estimation.

For 256-bit hash functions, each 256-bit (32 byte) hash value corresponds to 32 characters in ASCII encoding. However, for embedding quotable signatures into HTML code, the hash values need to be encoded as text. For compatibility, we investigate Base64 encoding. In Base64 encoding, each 256 bit string requires 44 characters which equals $44/4 = 11$ tokens. In other words, each hash value is worth eleven tokens of storage. Therefore, a sub-tree of height up to four may be better represented by words than a hash.

[3] 4.79 characters according to `https://norvig.com/mayzner.html`; 5.1 characters according to `https://www.wolframalpha.com/input/?i=average+word+length`.

Assuming "a single, continuous quote", worst case for verification path length is $2\lceil \log_2(n/2) \rceil$. So, for 24 tokens, the worst case signature equals the length of the text: $24 \cdot 4 = 96 = 2 \cdot \lceil \log_2(12) \rceil \cdot 11 + 2 \cdot 4$. Therefore, the minimum text size for a quotable signature is about 24 tokens or about 96 characters.

2.3 Signatures and Verification

We propose to use a standard signature scheme for signing the root node of the Merkle tree. In order to reduce the overhead of the overall scheme, we recommend to use an elliptic curve (ECC) signature scheme instead of RSA or DSA, for example `ecdsa_secp256r1_sha256` or `ed25519` from the TLS 1.3 draft. This gives an overhead of only 256 bit (32 bytes) for the signature.

For the verification of the root node, the verifier needs to know and trust the public key of the author. This can be achieved by a standard certification scheme.

3 Acknowledgements

The research reported in this paper has been supported in part by the German Research Foundation (DFG) within the project D.3 under RTG 2050 "Privacy and Trust for Mobile Users" and in part by the German Federal Ministry of Education and Research (BMBF) within the project "Scrutinise and Thwart Disinformation (DORIAN)".

Literatur

[Me79] Merkle, R. C.: Secrecy, authentication, and public key systems, Ph.D. thesis, Electrical Engineering, Stanford, 1979.

[RSL17] Ruchansky, N.; Seo, S.; Liu, Y.: CSI: A Hybrid Deep Model for Fake News Detection. In: CIKM. ACM, S. 797–806, 2017.

[SFR17] Singhania, S.; Fernandez, N.; Rao, S.: 3HAN: A Deep Neural Network for Fake News Detection. In: ICONIP (2). Bd. 10635. Lecture Notes in Computer Science, Springer, S. 572–581, 2017.

[Sh51] Shannon, C. E.: Prediction and entropy of printed English. The Bell System Technical Journal 30/1, S. 50–64, Jan. 1951.

IT-Grundschutz für die Container-Virtualisierung mit dem neuen BSI-Baustein SYS. 1.6

Christoph Haar[1], Erik Buchmann[2]

Abstract: Die Container-Virtualisierung baut auf eine komplexe IT-Landschaft auf, in der Hardware, Betriebssystem und Anwendungen von verschiedenen Parteien bereitgestellt und genutzt werden. Der IT-Sicherheit kommt daher eine große Bedeutung zu. Es gibt jedoch wenig Erfahrung mit der Absicherung der Container-Virtualisierung: Das Grundschutz-Kompendium und die Standards zur Risikoanalyse des Bundesamts für Sicherheit in der Informationstechnik (BSI) wurden erst im November 2017 in überarbeiteter Form neu eingeführt, und der BSI-Baustein SYS. 1.6 zur Container-Virtualisierung wurde erst im Mai 2018 als Community Draft veröffentlicht. Ziel dieser Arbeit ist die Erprobung des neuen Baustein SYS. 1.6 an einem konkreten Fallbeispiel. Dazu wenden wir den neuen Baustein auf ein typisches Docker Szenarion „Shop“ an und gehen die Gefährdungsanalyse, Docker-spezifische Gefährdungen sowie entsprechende Maßnahmen zur Abwendung dieser Gefährdungen ein. Wir haben festgestellt, dass der Baustein SYS. 1.6 des Grundschutz-Kompendiums eine umfassende Hilfestellung zur Absicherung der Container-Virtualisierung bietet, und in der Praxis gut anwendbar ist. Wir haben jedoch zwei zusätzliche Gefährdungen identifiziert, die der Baustein nicht ausreichend berücksichtigt.

Keywords: IT-Grundschutz; IT-Sicherheit; Container-Virtualisierung; Docker Container

1 Einleitung

Die Container-Virtualisierung ermöglicht es, innovative und cloudbasierte Anwendungen auf eine agile, kosteneffiziente Weise umzusetzen, auszuliefern und zu warten. Ein prominentes Beispiel ist das Open-Source Projekt Docker [Dob]. Container erlauben es, die eigentliche Anwendung von der IT-Infrastruktur zu trennen. So wird es beispielsweise möglich, ein vorkonfiguriertes Betriebssystem-Image zusammen mit einer Anwendung in einen Container zu packen und in einer Testumgebung zu prüfen. Der selbe Container lässt sich dann in das Produktivsystem übertragen und bei gewachsenem Ressourcenbedarf auf einen größeren Host-Rechner umziehen. Dafür benötigt die Container-Virtualisierung eine komplexe IT-Landschaft, in der verschiedene Parteien Softwarekomponenten oder Hardwareressourcen zur Verfügung stellen, Container bereitstellen oder die Virtualisierungsumgebung betreiben [Gö17]. Container können sensible Firmendaten oder personenbezogene Informationen enthalten. Unternehmen müssen daher bei Nutzung der Container-Virtualisierung ihr IT-Sicherheitskonzept überarbeiten.

[1] Hochschule für Telekommunikation Leipzig, haar@hft-leipzig.de
[2] Hochschule für Telekommunikation Leipzig, buchmann@hft-leipzig.de

doi:10.18420/inf2019_65

Wenn das Sicherheitskonzept auf dem IT-Grundschutz [Bu11] des Bundesamts für Sicherheit in der Informationstechnik (BSI) beruhen soll oder im Rahmen einer ISO 27001 Zertifizierung auf dem IT-Grundschutz aufgebaut [Bu14] wird, ist dies schwierig. Der Baustein SYS. 1.5 Virtualisierung des aktuellen BSI Grundschutz-Kompendiums [Bu19] zielt auf eine Hypervisor-Visualisierungsschicht ab. Im Mai 2018 wurde ein Community Draft für einen neuen Baustein SYS. 1.6 „Container“ [Bu] für die Container-Virtualisierung mittels Docker oder alternativer Technologien veröffentlicht. Dieser Baustein hat jedoch noch immer einen vorläufigen Charakter. Es gibt daher keine Erfahrungen, ob die darin beschriebenen Gefährdungen und Maßnahmen in der Praxis ausreichen, um ein gegebenes Anwendungsszenario ausreichend abzusichern. Wir haben eine Absicherung nach dem aktuellen IT-Grundschutz für ein typisches Container-Szenario durchgeführt:

Szenario: *Ein Einzelhändler verwendet einen Web-Shop, um sein Ladengeschäft zu ergänzen. Der Web-Shop verfügt über eine eigene Datenbank mit Produktbeschreibungen und Kundenkonten. Darüber hinaus ist der Web-Shop mit einem Internet-Zahlungssystem ausgestattet, das über einen Dienstleister Bezahlvorgänge über unterschiedliche Kanäle sicher abwickelt. Anwendung, Datenbank und Zahlungssystem sind auf verschiedene Docker-Container aufgeteilt. Die Container werden auf einem eigenen Rechner in einer On-Premise-Umgebung ausgeführt, bei der der Einzelhändler nicht nur für die Container verantwortlich ist, sondern auch die Infrastruktur und die Container-Plattform betreibt.*

In dieser Arbeit geben wir in geraffter Form unsere Erkenntnisse aus der Anwendung des IT-Grundschutzes wieder und zeigen anhand unseres Szenarios, dass der neue Baustein SYS.1.6 in der Praxis gut angewendet werden kann. Eine ausführliche Version ist als Preprint [HB18] verfügbar. Aufbauend auf [Ba17] und [BHB18] haben wir uns auf die Container-Virtualisierung konzentriert. Das heißt, wir haben für die bereits sehr gut untersuchte Absicherung [Dä18; Ec13] der Infrastruktur sowie der organisationsübergreifenden Aspekte ein bestehendes Sicherheitskonzept nach IT-Grundschutz vorausgesetzt. Wir haben nach BSI-Standard 200-2 [Bu17a] den Informationsverbund für unser Docker-System modelliert, dafür eine Schutzbedarfsfeststellung durchgeführt, und die in den BSI-Bausteinen SYS. 1.5 Virtualisierung und SYS. 1.6 Container beschriebenen Elementargefährdungen analysiert. Da einige Daten den Schutzbedarf „hoch“ erfordern, haben wir eine Risikoanalyse nach BSI-Standard 200-3 [Bu17b] zur Identifikation und Behandlung von zusätzlichen Gefährdungen für unser Docker-Szenario durchgeführt. Im Anschluss haben wir analysiert, inwiefern sich die dabei identifizierten Gefährdungen und Maßnahmen von denen des IT-Grundschutz-Kompendiums unterscheiden. Dabei hat sich gezeigt, dass der BSI-Baustein SYS. 1.6 das BSI-Grundschutz-Kompendium für die praktische Absicherung der Container-Virtualisierung gut anwendbar ist. Wir haben jedoch zwei zusätzliche Gefährdungen identifiziert, die vom Baustein nicht ausreichend berücksichtigt werden.

Aufbau der Arbeit: Abschnitt 2 beschreibt die Grundlagen dieser Arbeit. In Abschnitt 3 und 4 führen wir eine Risikoanalyse für Docker nach BSI-Standard durch und vergleichen unsere Erkenntnisse mit denen des BSI. In Abschnitt 5 verallgemeinern wir unsere Erkenntnisse. Die Arbeit schließt mit einer Zusammenfassung in Abschnitt 6.

2 Grundlagen

In diesem Abschnitt stellen wir das Docker-System, die BSI-Bausteine SYS. 1.5 und SYS. 1.6 [Bu] sowie die Vorgehensweisen zur Standardabsicherung und Risikoanalyse nach den aktuellen BSI-Standards [Bu17a; Bu17b] vor.

2.1 Docker-Container

Die Container-Virtualisierung hat sich aus der Hypervisor-Virtualisierung [Ch07] entwickelt. Ein Hypervisor zieht eine Abstraktionsschicht zwischen Host-System und den darauf ablaufenden Gast-Systemen ein. Dies hat unter anderem den Nachteil, dass für jeden Gast ein vollständiges Betriebssystem aufzusetzen ist. Im Gegensatz dazu werden bei der leichtgewichtigen Container-Virtualisierung Container zusammengestellt, die nur die Anwendung und eine leichtgewichtige Ablaufumgebung enthalten. Die Container nutzen also den Kernel des Host-Betriebssystems mit. Dies ermöglicht es, Systemressourcen wie Prozessor, Netzwerk oder Speicher effizient zu nutzen, und Applikationen über Systeme hinweg zu verschieben, ohne dabei komplette Betriebssysteme mit zu migrieren. Auf der anderen Seite wird es jedoch schwerer, mehrere Container, die auf dem selben Host-System ablaufen, zuverlässig voneinander zu isolieren.

Eine sehr häufig eingesetzte Lösung für die Container-Virtualisierung ist das auf einem Linux-Betriebssystem aufsetzende Docker. Linux-typisch besteht die Docker Architektur [Dob] aus Docker Client, Docker Daemon, Docker Registry und den Docker Objekten (Images, Docker Files, Container). Der Docker Client und Docker Daemon bilden zusammen die Docker Engine. Ein Container enthält zwei Hauptverzeichnisse: /bin enthält die Binärdateien und /lib die dynamischen Bibliotheken und Kernel-Module, die für die Funktionalität eines Containers benötigt werden. Client und Daemon können auf dem gleichen Host-System laufen oder der Client wird mit einem Remote Daemon verbunden. Die externe Kommunikation findet über eine REST API, ein UNIX Socket oder eine andere Netzwerkschnittstelle statt. Docker ist in seinen Grundeinstellungen so konfiguriert, dass nach Images aus dem Docker Hub gesucht wird. Es ist auch möglich, eine private Registry für Images anzulegen (Docker Trusted Registry).

Abbildung 1 beschreibt eine typische Docker-Installation, wie wir sie auch für unser Anwendungsszenario zugrunde gelegt haben: Auf dem Linux-Kernel setzt die Docker Engine auf. Für jede sachlogisch voneinander getrennte Aufgabe wird ein eigener Container betrieben. In unserem Fall sind dies drei Container, die voneinander isoliert eine Datenbank, ein Zahlungssystem und eine Web-Anwendung für den Online-Shop bereitstellen. Die Container kommunizieren über Linux-übliche Netzwerkschnittstellen miteinander und mit dem Internet (schwarze Pfeile). Zu diesem Zweck nutzen sie Funktionen der Docker Engine (graue Pfeile). Im Folgenden konzentrieren wir uns auf die Absicherung des in der Abbildung gestrichelt dargestellten Bereichs nach dem Ende 2017 überarbeiteten IT-Grundschutz. Eine Risikoanalyse nach dem alten IT-Grundschutz ist Teil unserer Vorarbeiten [BHB18].

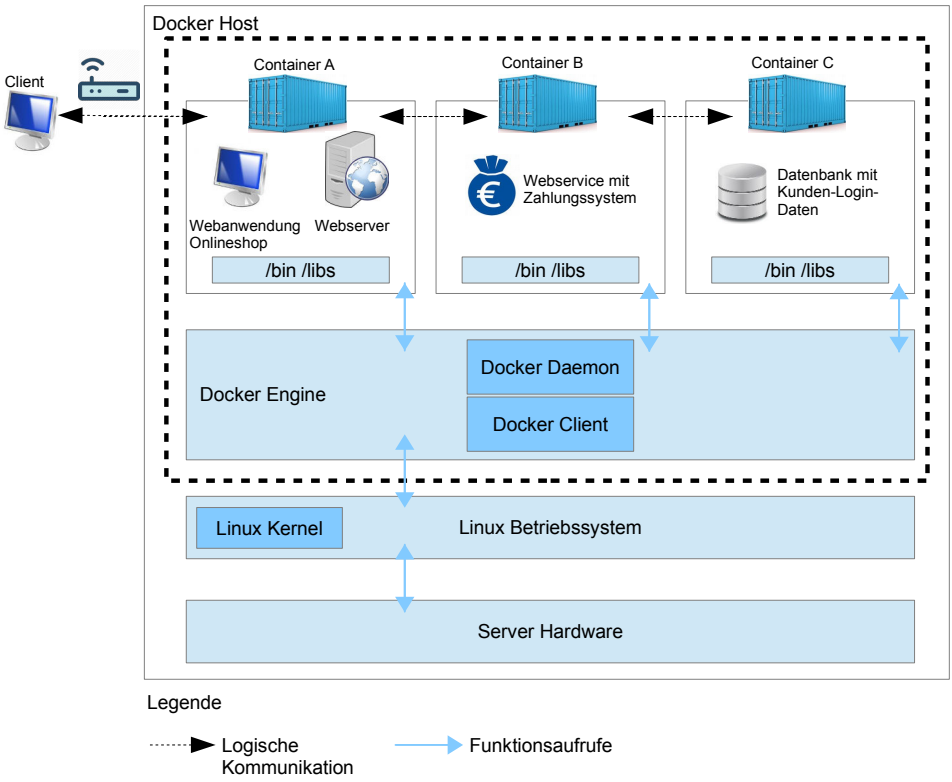


Abb. 1: Beispiel-System mit Dockerarchitektur

2.2 Standard-Absicherung und Risikoanalyse nach BSI

Der BSI-Standard 200-2 beschreibt, in welchen Schritten eine Standard-Absicherung eines Systems durchzuführen ist [Bu17a].

1. Zunächst ist der **Geltungsbereich** („Informationsverbund“) festzulegen, für den das Sicherheitskonzept realisiert werden soll. Der Geltungsbereich für unser Szenario ist in Abb. 1 gestrichelt dargestellt.
2. Bei der **Strukturanalyse** werden die Prozesse, Anwendungen, IT-Systeme, Infrastrukturen, etc. im Geltungsbereich aufgelistet.
3. Mit Hilfe der **Schutzbedarfsfeststellung** wird ein angemessener Schutz für die

Geschäftsprozesse, die darin verarbeiteten Informationen und die verwendete Informationstechnik ermittelt.
4. Bei der **Modellierung** werden Sicherheitsanforderungen und umzusetzende Maßnahmen mit den **Bausteinen** des IT-Grundschutz-Kompendiums [Bu19] identifiziert.
5. Mit dem **IT-Grundschutz-Check** wird geprüft, ob bereits umgesetzte Maßnahmen zur Absicherung des Informationsverbunds ein ausreichendes Schutzniveau bieten.
6. Wenn für ein Zielobjekt ein hoher oder sehr hoher Schutzbedarf besteht, kein passender BSI-Baustein existiert, oder das Zielobjekt auf eine Art und Weise betrieben wird, die der existierende Baustein nicht berücksichtigt, ist eine **Risikoanalyse** durchzuführen.

2.3 SYS. 1.5 Virtualisierung und SYS. 1.6 Container

Im BSI-Grundschutz-Kompendium [Bu19] werden abzusichernde Zielobjekte in Form von Bausteinen beschrieben. Für jedes Zielobjekt wird im Rahmen des Bausteins eine Zielstellung definiert, die das Ergebnis der Absicherung des Zielobjektes beschreibt. Darüber hinaus wird in jedem Baustein festgelegt, welche Bestandteile zur Absicherung des Zielobjektes zum Baustein gehören und welche nicht. Im Weiteren werden in jedem Baustein spezifische Gefährdungen für das Zielobjekt beschrieben. Zur Abwendung dieser Gefährdungen werden in jedem Baustein Anforderungen definiert. Zuletzt werden in jedem Baustein zusätzliche Informationen zu Gefährdungen und Sicherheitsmaßnahmen bereitgestellt.

Zur Absicherung des Docker-Systems benötigen wir zwei Bausteine. Der Virtualisierungs-Baustein SYS. 1.5 des BSI-Kompendiums [Bu19] behandelt die Gefährdungslage für Virtualisierungs-Systeme. Zwar adressiert der Baustein explizit nicht die Container-Virtualisierung. Da diese jedoch auf eine klassische Virtualisierung aufsetzt, haben wir SYS. 1.5 in unsere Analyse mit einbezogen. Der Baustein identifiziert folgende Gefahren:

- Fehlerhafte Planung der Virtualisierung
- Fehlerhafte Konfiguration der Virtualisierung
- Unzureichende Ressourcen für virtuelle IT-Systeme
- Informationsabfluss oder Ressourcen-Engpass durch Snapshots
- Ausfall des Verwaltungsservers für Virtualisierungs-Systeme
- Missbräuchliche Nutzung von Gastwerkzeugen
- Kompromittierung der Virtualisierungssoftware

Der im Mai 2018 als Community Draft veröffentlichte Baustein SYS. 1.6 [Bu] beschreibt folgende Gefährdungen für die Container-Virtualisierung:

- Schwachstellen in Images
- Administrative Zugänge ohne Absicherung
- Tool-basierte Orchestrierung ohne Absicherung
- Datenverluste durch fehlende Persistenz
- Vertraulichkeitsverlust von Zugangsdaten

Beide Bausteine sind hersteller- und produktneutral verfasst. Damit bleibt die Frage offen, ob die in den Bausteinen aufgeführten Elementargefährdungen und spezifischen Gefährdungen auch die für Docker spezifischen Bedrohungen für die IT-Sicherheit mit abdecken, und wie gut es möglich ist, diese Bedrohungen mit den Bausteinen als Handlungsunterstützung zu identifizieren und ihnen die passenden Maßnahmen gegenüberzustellen.

2.4 Identifikation zusätzlicher Gefährdungen

In einer Risikoanalyse sollen zusätzliche Gefährdungen identifiziert, eingeschätzt und um Maßnahmen ergänzt werden, die über die in den Bausteinen aufgeführten Gefährdungen hinausgehen. Zu diesem Zweck gibt der BSI-Standard 200-3 [Bu17b] folgenden Rahmen zur Ermittlung zusätzlicher Gefährdungen vor:

- Welche Gefahren aus dem Bereich „höhere Gewalt“ sind besonders relevant?
- Bestehen organisatorische Mängel, die die Informationssicherheit beeinträchtigen?
- Kann die Sicherheit durch menschliche Fehlhandlungen beeinträchtigt werden?
- Welche speziellen Sicherheitsprobleme kann technisches Versagen hervorrufen?
- Welche Gefährdungen können durch externe Angriffe entstehen?
- Ist es Mitarbeitern möglich, den Betrieb des Zielobjekts mutwillig zu beeinträchtigen?
- Können von Objekten außerhalb des Informationsverbunds Gefahren ausgehen?
- Welche Hinweise geben Herstellerdokumentationen sowie Informationen von Dritten?

Diese Fragen sollen von Experten, Mitarbeitern, Administratoren und Benutzern gemeinsam bearbeitet werden.

3 Schutzbedarfsfeststellung und Elementargefährdungen für Docker

In diesem Abschnitt entwickeln wir nach BSI-Standard 200-2 Abschnitt 8 [Bu17a] eine Standard-Absicherung für unser Anwendungsszenario. Wir beginnen mit der Modellierung des Informationsverbunds und einer Schutzbedarfsfeststellung. In einem nächsten Schritt untersuchen wir die in den Bausteinen SYS. 1.5 und SYS. 1.6 vorgegebenen Elementargefährdungen auf ihre Anwendbarkeit für Docker.

3.1 Das Docker-System

Unser in Abbidlung 1 dargestelltes Docker-Szenario besteht aus einem Online-Shop, dessen Komponenten auf drei Container aufgeteilt sind. In Container A läuft eine Webanwendung auf einem Webserver, die ein Shop-System incl. Einkaufskorb, Kundenrezensionen etc. umsetzt. In Container B wird ein Webservice betrieben, der die Zahlungsabwicklung für unseren Onlineshop realisiert. Container C enthält eine Datenbank, die Produkt-, Kunden-

und Bestelldaten beinhaltet. Diese drei Container werden isoliert voneinander betrieben und bilden zusammen mit der Docker Engine das Host-System. Der Informationsverbund ist nachfolgend zusammengefasst:

Nr.	Datenobjekt	Beschreibung
D1	Personendaten	Einzelangaben zu einer natürlichen Person
D2	Nutzdaten	Fachdaten der Anwendungen und Services
D3	Accountdaten	Anmelde- und Berechtigungsdaten der Anwender
D4	Konfigurationsdaten	Daten zur Änderung, Einstellung und Anpassung
D5	Protokolldaten	Statusinformationen und Funktionen

Nr.	Beschreibung	verarbeitete Daten	Software
A1	Webanwendung	D1, D2, D3, D4, D5	Allgemeine Anwendung z.B. PHP
A2	Webserver	D1, D2, D4, D5	Apache Webserver
A3	Webservice	D2, D3, D4, D5	REST-basierter Dienst
A4	Datenbank	D1, D2, D3, D4, D5	Allgemeine Datenbank z.B. MySQL

Nr.	Beschreibung	verarbeitete Daten	IT-System
SSW1	Docker Software	D4, D5	S1 und S1

Nr.	Beschreibung	verarbeitete Daten	Plattform	Ort
S1	Host-System	D1, D2, D3, D4, D5	x86 Linux-Server	RZ 1

Dieser Informationsverbund ist typisch für viele Docker-Installationen. Da wir uns auf die Absicherung von Docker konzentrieren wollen, haben wir unter S1 „Host-System“ die gesamte Host-Umgebung zusammengefasst, d.h., das Rechenzentrum mit dem Host-Rechner und dem darauf installierten Host-Betriebssystem.

3.2 Schutzbedarfsfeststellung

Um herauszufinden, welche Maßnahmen für den Schutz der Objekte in unserm Informationsverbund angemessen sind, haben wir eine Schutzbedarfsfeststellung durchgeführt. Wir verwenden die im Standard 200-2 [Bu17a] definierten Schutzbedarfskategorien „normal“, „hoch“ und „sehr hoch“.

Bei der Schutzbedarfsfeststellung vererben sich die Schutzbedarfe einzelner Datenobjekte (D1-D4) auf die Anwendungen (A1-A4), die diese Daten verarbeiten, und von dort auf die Systeme (S1), auf denen diese Anwendungen ablaufen. Speichert ein System Daten mit unterschiedlichen Schutzbedarfen, so wird dem System der höchste dieser Schutzbedarfe zugewiesen.

Daraus ergibt sich eine Besonderheit für die Container-Virtualisierung: Sämtliche Container laufen möglicherweise auf der gleichen physischen Maschine (S1). Funktionen des Betriebssystem-Kernels des Hosts werden von allen Containern gleichermaßen verwendet. Zudem funktioniert das Gesamtsystem – in unserem Falle der Online-Shop – nur, wenn sämtliche Container betriebsbereit sind. Deswegen vererbt sich der höchste Schutzbedarf jedes einzelnen Containers automatisch auf den gesamten Informationsverbund. Für die Schutzbedarfsfeststellung genügt es deswegen, über alle Container hinweg nach den Daten oder Diensten mit dem höchsten Schutzbedarf bezüglich Vertraulichkeit, Integrität und Verfügbarkeit zu suchen und diesen dann für das Gesamtsystem zu übernehmen. Für unser Anwendungsszenario bedeutet dies:

- **Vertraulichkeit:** In Container C wird eine Datenbank betrieben, die Kundendaten mit Personenbezug speichert. Daher besteht für den Informationsverbund ein hoher Schutzbedarf für den Grundwert Vertraulichkeit.
- **Integrität:** In Container B werden die Zahlungsvorgänge der Kunden abgewickelt. Der Schutzbedarf des Informationsverbunds bezüglich der Integrität ist deshalb hoch.
- **Verfügbarkeit:** Der Web-Shop ist geschäftskritisch, funktioniert aber nur, wenn alle drei Container sowie das Betriebssystem und die Hardware verfügbar sind. Deswegen besteht für den gesamten Informationsverbund ein hoher Schutzbedarf für die Verfügbarkeit.

Unsere zentrale Erkenntnis aus der Schutzbedarfsfeststellung ist, dass die Schutzbedarfe für Vertraulichkeit, Integrität und Verfügbarkeit für typische Einsatzszenarien mindestens „hoch“ sind. Dies gilt beispielsweise für alle von Docker aufgeführten Kundenprojekte [Doa]. Bezogen auf den BSI-Grundschutz bedeutet dies, dass in jedem Fall nach der Standard-Absicherung eine Risikoanalyse durchzuführen ist (s. Abschnitt 4).

3.3 Analyse der Elementargefährdungen

Nach der Schutzbedarfsfeststellung sieht das BSI die Modellierung eines Grundschutzkonzepts auf der Basis der im Grundschutz-Kompendium definierten Bausteine vor. Der erste Schritt besteht dabei in der Prüfung der in den Bausteinen genannten Elementargefährdungen. Wir haben bereits festgestellt, dass für das Docker-System die Bausteine SYS. 1.5 und SYS. 1.6 relevant sind. Zusammen listen die beiden Bausteine 25 Elementargefährdungen auf. Für eine detaillierte Auseinandersetzung mit Elementargefährdungen wie Datenverlust oder Ressourcenmangel im Docker-Informationsverbund verweisen wir auf den Preprint [HB18] dieser Arbeit. Im nächsten Abschnitt legen wir unseren Fokus auf den Umgang mit Docker-spezifischen Bedrohungen, die über Elementargefährdungen hinausgehen.

4 Docker-spezifische Gefährdungen

Da in unserem Informationsverbund mehrere Objekte einen über „normal“ hinausgehenden Schutzbedarf ausweisen, ist eine Riskioanalyse zur Identifikation und bewertung zusätzlicher Gefährdungen erforderlich, gefolgt von einer Analyse der Risikobehandlungsoptionen.

4.1 Identifikation zusätzlicher Gefährdungen

Gemeinsam mit Experten der Open Telekom Cloud haben wir eine Risikoanalyse durchgeführt (vgl. Abs. 2.2). Dabei haben wir 14 Gefährdungen identifiziert (s. Abbildung 2). 12 dieser Gefährdungen sind auch in den Bausteinen SYS. 1.5 und SYS. 1.6 als spezifische Gefährdungen enthalten. Darüber hinaus konnten wir zwei zusätzliche Gefährdungen identifizieren (gestrichelt in Abb. 2 dargestellt).

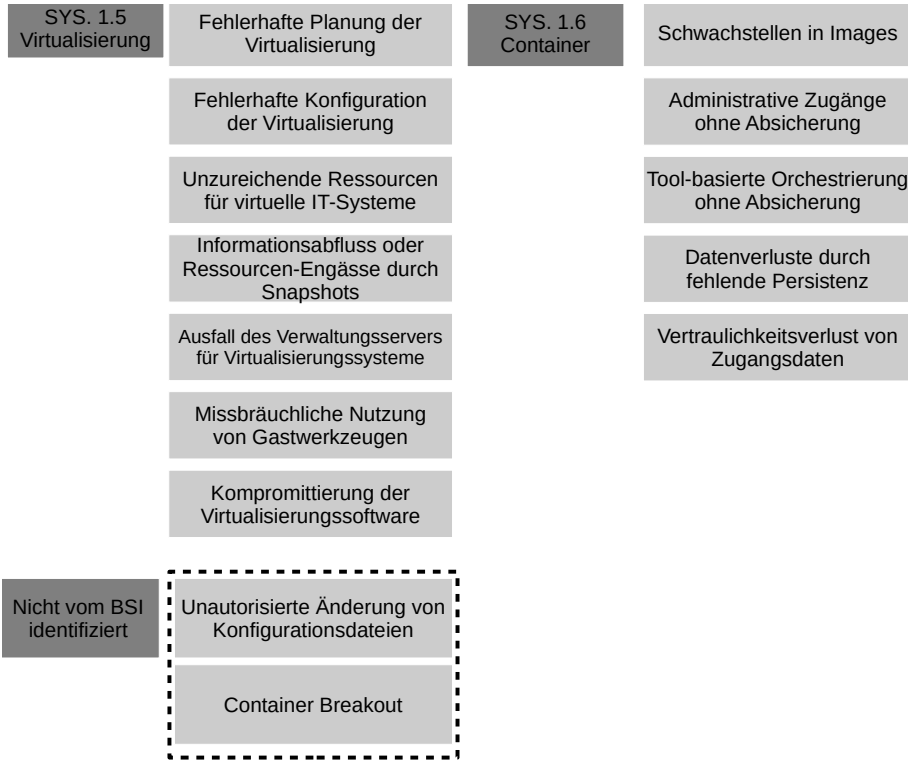


Abb. 2: Spezifische Gefährdungen für das Docker-System

Ein **Container Breakout** [Ro] wird möglich, wenn die Container durch Schwachstellen in der Implementierung nicht lückenlos voneinander isoliert sind. Bei einem erfolgreichen Breakout kann der Angreifer mit den Privilegien des Containers, aus dem er ausgebrochen ist,

auf Daten oder Dienste des Host-Systems oder anderer Container zugreifen. Ein Container Breakout beeinträchtigt nicht nur die Vertraulichkeit, sondern auch die Verfügbarkeit und die Integrität von anderen Objekten im Informationsverbund. Der Baustein SYS. 1.6 adressiert diesen Punkt nur indirekt durch die Basis-Anforderung SYS. 1.6. A2 zur Planung eines Netzzonenkonzepts.

Durch **unautorisierte Änderungen an Konfigurationsdateien** der virtuellen Infrastruktur können erhebliche und tiefgreifende Schäden entstehen, ebenso wie durch vorsätzliche oder versehentliche Fehlkonfigurationen der Netzzuordnung. Hier stellt insbesondere der Docker Daemon eine Angriffsoberfläche dar, da dieser root-Privilegien besitzt und die Funktionsfähigkeit aller Container beeinflussen kann. Für Vertraulichkeit, Integrität oder Verfügbarkeit der Objekte im Informationsverbund ist die Integrität von Konfigurationsdaten daher ausschlaggebend. Auch diese Gefährdung wird nur mittelbar durch die Standard-Anforderungen A17 in SYS. 1.5 (Netzzuordnungen im Virtualisierungslayer) und A12 in SYS. 1.6 (Freigabe von Images) adressiert. Details zu allen von uns identifizierten Gefährdungen finden sich in [HB18].

4.2 Risikoeinstufung und Risikobewertung

Im nächsten Schritt muss das Risiko ermittelt werden, welches von der jeweiligen Gefährdung ausgeht. Dazu wird nach BSI-Standard 200-3 [Bu17b] eine qualitative Risikobewertung herangezogen. Unsere zentrale Erkenntnis ist hier, dass sich alle spezifischen Gefährdungen für das Docker-System auf technische Schwachstellen beziehen, für die sich Angriffe automatisieren lassen. Wird beispielsweise ein Exploit bekannt, durch den sich ein Container Breakout durchführen lässt, so kann dieser Exploit auch automatisiert auf eine große Zahl von anfälligen Containern angewendet werden. Wir gehen aus diesem Grund davon aus, dass die Eintrittshäufigkeit für jede Gefährdung für Docker „sehr häufig“ ist. Für die Risikobewertung nach BSI-Standard 200-3 genügt es also, die Schadenshöhen der Gefährdungen zu ermitteln.

Im Folgenden stellen wir die Risikobewertung für die Gefährdung „Container Breakout“ beispielhaft dar. Für die „Unautorisierte Änderung von Konfigurationsdateien“ gelten die gleichen Eintrittshäufigkeiten, Auswirkungen und Risiken wie für den Container Breakout. Es kommen noch die Beeinträchtigung der Integrität und Verfügbarkeit hinzu.

Docker-System	Vertraulichkeit: hoch Integrität: hoch Verfügbarkeit: hoch	
Gefährdung **Container Breakout**	Beeinträchtigte Grundwerte: Vertraulichkeit	
Eintrittshäufigkeit ohne zus. Maßnahme: sehr häufig	**Auswirkungen** ohne zusätzliche Maßnahmen: beträchtlich	**Risiko** ohne zusätzliche Maßnahme: hoch
Beschreibung: Ein Gefährdungsszenario ist der Container Breakout, der einem Angreifer Zugriff auf das Host-System oder auf weitere Container im gleichen System erlaubt, und zwar mit den Privilegien des Containers, aus dem der Ausbruch erfolgte.		
Bewertung: Ein Container Breakout würde zum Verlust der Vertraulichkeit von z.B. Kundendaten führen. Diese gelten als hoch schutzbedürftig, sodass das Schadensausmaß bei einem Container Breakout beträchtlich und das Risiko als hoch einzustufen ist.		

4.3 Risikobehandlung

Grundsätzlich stehen als Risikobehandlungsoptionen die Risikoreduktion durch zusätzliche Maßnahmen oder durch Umstrukturierung der Prozesse, der Risikotransfer oder die Risikoakzeptanz zur Verfügung [Bu17b]. Aufgrund der Risikoeinstufung „hoch“ scheidet die Risikoakzeptanz aus. Risikotransfer oder Umstrukturierung liegen außerhalb unseres Geltungsbereichs. Im Folgenden gehen wir auf die Risikoreduktion durch zusätzliche Maßnahmen für die von uns identifizierten zusätzlichen Gefährdungen ein. Für die Gefährdungen, die auch das BSI identifiziert hat, verweisen wir auf die Bausteine SYS.1.5 und SYS.1.6.

Container Breakout (Risikokategorie: hoch)

Definition der Systembenutzer Docker-Container sind nicht als privilegierte Container zu betreiben, damit Angreifer im Erfolgsfall nur unprivilegierten Zugriff auf andere Ressourcen erhalten.

Rechtemanagement Es sind die Berechtigungen für alle definierten Benutzergruppen auf Minimalität zu prüfen.

Rollenaufteilung Es ist auch für virtuelle IT-Systeme eine Aufteilung in verschiedene Rollen notwendig. Linux-Schutzmaßnahmen wie apparmor, selinux, seccomp, Filter und namespaces auf dem Host-System können das Risiko eines Ausbruchs reduzieren.

Unautorisierte Änderung von Konfigurationsdateien (Risikokategorie: hoch)

Prüfsummen Die Prüfung auf unautorisierte Änderungen der Konfigurationsdateien kann beispielsweise mittels Werkzeugen wie OS-SEC erfolgen [OS].

Docker Bench for Security Docker ab Version 1.10.0 bietet das Docker Bench for Security Script [Ce] an, welches die eigene Docker Konfiguration prüft.

Konfiguration der Netzfunktionen Bekannte Linux-Werkzeuge wie beispielsweise Puppet [AJ17] können die Netzkomponenten zentral überwachen.

Benennung virtueller Netze Eine aussagekräftige Benennung der Netze anhand ihrer Funktion vermeidet ein versehentliches Verbinden mit dem falschen Netzwerk [AJ17].

Speicher-Zentralisierung Wenn ein Dateiverzeichnis des Containers mit dem Host- System verknüpft wird, muss dieses die Isolation von Betriebssystem, Systembibliotheken [Va17] und gemeinsamen Anwendungen sicherstellen.

Monitoring Das Monitoring lässt sich durch den Einsatz eines Linux-Servers mit den systemeigenen Monitoring Systemen wie Nagios bewerkstelligen [AJ17].

Kommunikation zwischen Containern Bei aktivem Container Linking [Ja] müssen Container, die nicht miteinander kommunizieren dürfen, auf separaten Hosts ablaufen.

In einem letzten Schritt müssen nun die von uns vorgeschlagenen Maßnahmen mit den Anforderungen der bestehenden BSI-Bausteine konsolidiert werden. Beispielsweise findet sich die von uns vorgeschlagene Definition der Systembenutzer als Maßnahme gegen den Container Breakout im Baustein SYS. 1.6 in der Anforderung A17 wieder. Andere Maßnahmen wie die Berücksichtigung der Kommunikation zwischen Containern finden sich in den BSI-Bausteinen nicht wieder. Aus Platzgründen verweisen wir für eine vollständige Übersicht auf unser Preprint [HB18].

5 Diskussion

In diesem Abschnitt diskutieren wir, inwiefern sich unsere Erkenntnisse auf die Container-Virtualisierung insgesamt sowie auf andere Anwendungsszenarien verallgemeinern lassen.

Container-Virtualisierung Der BSI-Baustein SYS. 1.6 ist bereits von der eingesetzten Technologie unabhängig definiert. Die von uns identifizierten zusätzlichen Gefährdungen sind jedoch ebenfalls nicht Docker-spezifisch. Im Gegensatz zur traditionellen Hypervisor-Virtualisierung [Ch07] nutzen die leichtgewichtigen Container Funktionen aus dem Kernel Host-Betriebssystem [Dob], beispielsweise um zu kommunizieren oder um Ressourcen zu allokieren. Diese Funktionen öffnen potentielle Zugriffspfade für einen Angreifer, um aus der isolierten Container-Umgebung auszubrechen. Auch unautorisierte Änderungen der Konfigurationsdateien stellen für Container eine Gefährdung dar. Jeder Container wird entsprechend seiner benötigten Berechtigungen konfiguriert. Unautorisierte Änderungen an den Konfigurationsdateien können daher einen erheblichen Einfluss auf die Integrität, Verfügbarkeit und Vertraulichkeit sowohl der Container als auch des Host-Systems haben. Es würde sich daher anbieten, diese beiden Gefährdungen explizit in den neuen Container-Bausteins SYS. 1.6 aufzunehmen.

Allgemeine Anwendungsszenarien Unsere Risikoanalyse hat ergeben, dass sich die von uns identifizierten zusätzlichen Gefährdungen für automatisierbare Angriffe eignen, sobald eine entsprechende Schwachstelle für eine Container-Technologie entdeckt wird.

Daher sind unsere Erkenntnisse über unser Anwendungsszenario und dessen konkrete Schutzbedarfe hinaus wichtig. Wir haben unsere Risikoanalyse auf der Basis einer Schutzbedarfsfeststellung durchgeführt, bei der die Bedarfe für Vertraulichkeit, Integrität und Verfügbarkeit für das Gesamtsystem mit „hoch“ festgesetzt wurde. Wir haben festgestellt, dass dies aufgrund des Maximumprinzips typisch ist für viele kommerzielle Anwendungen der Container-Virtualisierung. Für Anwendungsfälle, bei denen die Schadensauswirkungen ein „existenziell bedrohliches, katastrophales Ausmaß erreichen“ [Bu17a] können, ist jedoch eine umfassendere Risikoanalyse erforderlich. Ein Beispiel für so ein Anwendungsszenario könnte ein Krankenhaus sein, das medizinische Geräte über eine Container-Lösung steuert.

6 Zusammenfassung

Das Ziel dieser Arbeit bestand darin, zu untersuchen, wie gut die aktuellen BSI-Standards und der neue Baustein SYS. 1.6 auf einen typischen Anwendungsfall der Container-Virtualisierung angewendet werden können. Dazu haben wir eine Standard-Absicherung und eine Risikoanalyse nach BSI IT-Grundschutz für Docker Container in einer On-Premise-Umgebung durchgeführt. Wir haben festgestellt, dass der neue Baustein SYS. 1.6 in Verbindung mit dem Virtualisierungs-Baustein SYS. 1.5 ein wertvolles Werkzeug bei der Erstellung eines Sicherheitskonzepts für Docker darstellt. In unserem konkreten Anwendungsfall hat sich jedoch gezeigt, dass zwei zusätzliche Gefährdungen für Docker existieren, die im Rahmen des neuen Bausteins SYS. 1.6 noch nicht berücksichtigt wurden. Wir haben gezeigt, dass sich unsere gewonnenen Erkenntnisse nicht nur auf Docker-Szenarien beschränken sondern im Allgemeinen für Container-Technologien gelten. Daher ist eine Ergänzung des Baustein SYS. 1.6 um unsere zusätzlichen Gefährdungen sowie der dazugehörigen Maßnahmen und Anforderungen zu überlegen.

Literatur

[AJ17] Atug, M.; Jedecke, D.: iX Kompakt - Container und Virtualisierung. Heise Medien, 2017.

[Ba17] Bauer, S.: Erarbeitung eines Informationssicherheitskonzepts nach IT-Grundschutz für Docker Container. Bachelor-Arbeit, Hochschule für Telekommunikation Leipzig, Kopie s. http://www.webcitation.org/6xAkE4g1l/, 2017.

[BHB18] Buchmann, E.; Hartmann, A.; Bauer, S.: Informationssicherheitskonzept nach IT-Grundschutz für Containervirtualisierung in der Cloud. SICHERHEIT 2018/, 2018.

[Bu] Bundesamt für Sicherheit in der Informationstechnik: SYS.1.6 Container, https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Grundschutz/IT-Grundschutz-Modernisierung/BS_Container.html, abgerufen Sept. 2018.

[Bu11] Bundesamt für Sicherheit in der Informationstechnik: Webkurs IT-Grundschutz, IT -Grundschutz im Selbststudium. https://www.bsi.bund.de/, 2011.

[Bu14] Bundesamt für Sicherheit in der Informationstechnik: Zertifizierung nach ISO 27001 auf der Basis von IT-Grundschutz. https://www.bsi.bund.de/, 2014.

[Bu17a] Bundesamt für Sicherheit in der Informationstechnik: BSI-Standard 200-2, IT-Grundschutz-Methodik. https://www.bsi.bund.de/, 2017.

[Bu17b] Bundesamt für Sicherheit in der Informationstechnik: BSI-Standard 200-3, Risikomanagement. https://www.bsi.bund.de/, 2017.

[Bu19] Bundesamt für Sicherheit in der Informationstechnik: IT-Grundschutz-Kompendium - Edition 2019. https://www.bsi.bund.de/, 2019.

[Ce] Center for Internet Security: Docker Community Edition Benchmark, https://www.cisecurity.org, abgerufen Sept. 2018.

[Ch07] Chisnall, D.: The Definitive Guide to the Xen Hypervisor. Prentice Hall, 2007.

[Dä18] Dännart, S.; Diefenbach, T.; Hofmeier, M.; Rieb, A.; Lechner, U.: IT-Sicherheit in Kritischen Infrastrukturen–eine Fallstudien-basierte Analyse von Praxisbeispielen. Multi-Konferenz Wirtschaftsinformatik (MKWI'18)/, 2018.

[Doa] Docker Inc.: Docker Customers, https://www.docker.com/customers, abgerufen Sept. 2018.

[Dob] Docker Inc.: Docker Overview, https://docs.docker.com/engine/docker-overview, abgerufen Sept. 2018.

[Ec13] Eckert, C.: IT-Sicherheit: Konzepte-Verfahren-Protokolle. Walter de Gruyter, 2013.

[Gö17] Göbel, L.: Container-as-a-Service - Die Zukunft der Virtualisierung, https://www.cloudcomputing-insider.de/container-as-a-service-die-zukunft-der-virtualisierung-a-576244, abgerufen Sept. 2018, 2017.

[HB18] Haar, C.; Buchmann, E.: IT-Grundschutz für die Container-Virtualisierung mit dem neuen BSI-Baustein SYS. 1.6. In. Quality Content Of Saxony, 2018.

[Ja] Jacqueline von Ogden: The Top 5 Security Risks in Docker Container Deployment, https://www.cimcor.com/blog/the-top-5-security-risks-in-docker-container-deployment, abgerufen Sept. 2018.

[OS] OSSEC Project Team: OSSEC's Documentation, https://ossec-docs.readthedocs.io/en/latest, abgerufen Sept. 2018.

[Ro] Rob Shapland: Eine Schwachstelle in Container-Techniken erlaubt Angriffe auf den Host, https://www.searchsecurity.de/tipp/Eine-Schwachstelle-in-Container-Techniken-erlaubt-Angriffe-auf-den-Host, abgerufen Sept. 2018.

[Va17] Vasily Tarasov, L. R.: In Search of the Ideal Storage Configuration for Docker Containers. IEEE 2nd International Workshops on Foundations and Applications of Self* Systems (FAS*W)/, 2017.

Ende-zu-Ende-Sicherheit für die multimodale Mobilität in einer Smart City

Erik Buchmann[1], Franziska Plate[2]

Abstract: Im Zuge einer Mobilitätswende werden Konzepte der multimodalen Mobilität immer wichtiger. Multimodale Mobilität bedeutet, dass dem Nutzer in Abhängigkeit von persönlichen und externen Faktoren eine Kombination aus Reisemitteln angeboten, gebucht und abgerechnet wird, die sein Mobilitätsbedürfnis erfüllen. Zu den persönlichen Faktoren zählen Präferenzen wie Preis, Komfort oder Reisezeit, zu den externen die Verfügbarkeit von Verkehrsmitteln, Staus oder Umweltparameter. Dies erfordert eine komplexe Vernetzung von Verkehrsmitteln, Umweltsensoren, Mobilitäts- und Abrechnungsdienstleistern, intelligenten Verfahren zur Stau- und Klimavorhersage, sowie eine Echtzeitüberwachung der Nutzerposition. Der IT-Sicherheit kommt deswegen eine entscheidende Bedeutung zu. Wir untersuchen, inwieweit sich die multimodale Mobilität für den Nahverkehr in einem typischen Smart City-Szenario technisch absichern lässt. In Anlehnung an den IT-Grundschutz modellieren wir die Datenflüsse, die für die Umsetzung der multimodalen Mobilität erforderlich sind. Wir untersuchen, inwiefern die derzeit verfügbaren Konzepte der IT-Sicherheit für diesen Anwendungsfall geeignet sind, und führen eine Risikoanalyse durch. Unsere Arbeit zeigt, dass bei einer konsequenten Realisierung eines Sicherheitskonzepts das größte Risiko durch Fehlbedienung oder Fehlkonfiguration des Smartphones des Nutzers entsteht.

Keywords: IT-Sicherheit, Smart City, Multimodale Mobilität

1 Einleitung

Die Ausgestaltung der urbanen Mobilität wird gerade für Ballungsräume immer wichtiger. Dabei werden integrierte Mobilitätsangebote untersucht, die schienen- oder straßengebundene öffentliche Verkehrsmittel, Car-Sharing, Car-Pooling oder Leihfahrräder intelligent zu einer multimodalen Mobilitätsform zusammenführen, welche das Mobilitätsbedürfnis des Nutzers erfüllt. Dabei sind persönliche Präferenzen wie Preis, Komfort, Reisezeit oder Umweltfreundlichkeit zu berücksichtigen und externe Faktoren wie die Verfügbarkeit von Verkehrsmitteln, Verspätungen oder Umweltparameter mit einzubeziehen [GB14].

Die multimodale Mobilität erfordert eine komplexe Vernetzung von Verkehrsmitteln, Umweltsensoren, Mobilitäts- und Abrechnungsdienstleistern, intelligenten Verfahren zur Stau- und Klimavorhersage, sowie eine Echtzeitüberwachung der Nutzerposition. Sie ist daher eine typische Anwendung für eine Smart City-Plattform. Dabei übernimmt die Plattform

[1] Hft-Leipzig, Gustav-Freytag-Straße 43-45, 04277 Leipzig, buchmann@hft-leipzig.de
[2] Detecon International GmbH, Sternengasse 14 - 16, 50676 Köln, Franziska.Plate@detecon.com

doi:10.18420/inf2019_66

komplexe Funktionen entlang der Wertschöpfungskette der multimodalen Mobilität, von der Reiseplanung des Nutzers bis zur Abrechnung der tatsächlich in Anspruch genommenen Mobilitätsdienste. Die dafür erforderlichen Daten stammen aus dem öffentlichen Internet, Smartphones der Nutzer oder IoT-Komponenten. Die Akzeptanz einer Lösung für die multimodale Mobilität hängt nicht nur von Fragen der Vertraulichkeit und des Datenschutzes ab, sondern auch von der täglichen Verfügbarkeit des Dienstes. Spätestens bei der Abrechnung ist auch die Datenintegrität wesentlich. Der IT-Sicherheit [Ec18] kommt deswegen eine entscheidende Bedeutung zu [Bo19].

In dieser Arbeit analysieren wir, wie sich die multimodalen Mobilität in einem typischen Smart City-Szenario technisch absichern lässt. Dabei konzentrieren wir uns auf die Verknüpfung von Nahverkehrsmitteln. In unserem Fokus stehen nicht Schwachstellen existierender Implementierungen. Vielmehr untersuchen wir die Sicherheitsrisiken innerhalb der Wertschöpfungskette. In Anlehnung an den IT-Grundschutz des Bundesamts für Sicherheit in der Informationstechnik (BSI) modellieren wir die Datenflüsse und Übertragungswege, die für die multimodale Mobilität erforderlich sind. Wir untersuchen, ob derzeit verfügbaren Konzepte für diesen Anwendungsfall geeignet sind, und führen eine Risikoanalyse durch. Unser Ziel ist eine Ende-zu-Ende (E2E) Absicherung der Systeme und Übertragungswege entlang der Wertschöpfungskette. Unsere Arbeit zeigt, dass bei einer konsequenten Absicherung nach dem Stand der Technik das größte Risiko durch Fehlbedienung oder Fehlkonfiguration des Smartphones des Nutzers entsteht, und wir zeigen auf, um welche Risiken es sich handelt. Aus Platzgründen können wir hier nur eine Übersicht über unsere Erkenntnisse bieten. Details stehen in einem Arbeitsbericht [PB19] zur Verfügung.

In Abschnitt 2 beschreiben wir verwandte Arbeiten. In Abschnitt 3 führen wir eine Risikoanalyse für die multimodale Mobilität in einem Smart City-Szenario durch, gefolgt von Risikobehandlungsoptionen in Abschnitt 4. Wir schließen mit einem Fazit in Abschnitt 5.

2 Verwandte Arbeiten

Im Folgenden führen wir die multimodalen Mobilität, deren technische Grundlagen und Ansätze zur Absicherung ein. Wir setzen Grundkenntnisse zum IT-Grundschutz [Bu19] und zu Kommunikationsprotokollen wie WLAN oder LTE voraus.

2.1 Konzepte für die multimodale Mobilität

Die multimodale Mobilität ist ein Konzept, bei dem unterschiedliche Verkehrsmittel innerhalb einer Reiseroute miteinander kombiniert werden. Dabei umfasst das Konzept sowohl den Nahverkehr, wie z.B. Bike-Sharing, Car-Sharing, Straßenbahnen, Busse und Taxen, als auch den Fernverkehr, u.a. Züge, Flugzeuge und Schiffe. Nutzer können entweder eigenständig auf die verschiedenen Verkehrsmittel zurückgreifen, oder sie nutzen einen Planungsdienst. Bereits seit 2001 verfolgt die Deutsche Bahn diese Mobilitätsstrategie, welche

den Nutzer mit entsprechenden Konzepten von Haustür zu Haustür bringen soll [Ma06]. Auch über den Karten-Dienst Google Maps können Routen multimodal geplant werden. Dafür muss der Routenplanung bekannt sein, welche Verkehrsmittel wie, wann und wo zur Verfügung stehen. Ebenfalls muss die Routenplanung wissen, wo mögliche Umsteige-Punkte zwischen den Verkehrsmitteln liegen. Die Nutzung von unterschiedlichen Verkehrsmitteln innerhalb einer Route kann über ein umfassendes E-Ticket gelöst werden [JS14]. Hierbei kann das Fahrzeug mit dem selben Ticket entsperrt, genutzt und abgerechnet werden. Die Abrechnung erfolgt auf Basis der Nutzung und stellt eine erhebliche Anforderungen an Datenschutz und Datensicherheit [Ec18].

2.2 Technische Grundlagen der multimodalen Mobilität

Im Internet of Things (IoT) [Xi12] werden physische Gegenstände lesbar, erkennbar, auffindbar, adressierbar und/oder steuerbar. IoT-Geräte erhalten Sensorik sowie einen kleinen Prozessor und Speicher, wodurch sie kontextbezogene Entscheidungen [PP+16] treffen können. Dabei können sie über eine Kommunikationsverbindung auf Daten von anderen Geräten zugreifen. IoT-Konzepte sind daher bei der Realisierung von Smart City-Anwendungsfällen unverzichtbar, beispielsweise als Umweltsensor oder zur Überwachung und Abrechnung von Verkehrsmitteln. Gleichwohl nimmt die Angriffsfläche in der IoT-Umgebung aufgrund der Heterogenität von Geräten, Kommunikationsmedien, Anwendungen und Diensten vielfältig zu [HL17], während Sicherheitsmechanismen im IoT häufig vernachlässigt werden [Bu18].

Eine Smart City zielt auf die vollständige Vernetzung aller digitalen Anwendungsfälle über eine zentrale Smart City-Plattform ab. Diese ermöglicht den Informationsaustausch zwischen den einzelnen IoT-Komponenten sowie die Steuerung und Überwachung der eingesetzten IoT-Geräte. Wie eine solche zentrale Plattform aussehen kann, zeigt Abb. 1. Eine Connectivity Management-, Solution Enabling- und eine Big Data-Plattform bilden die Smart City-Plattform ab. Innerhalb der IoT-Anwendungsfälle kann es sowohl IoT-Geräte mit als auch ohne SIM-Karte geben. Die Komponenten, welche eine SIM-Karte besitzen, können ihre Daten über das Mobilfunknetz versenden. Dazu erhalten sie eine private Adresse, die über einen Access Point Name-Dienst (APN) auf eine öffentlich sichtbare IP-Adresse übertragen wird. Komponenten ohne SIM-Karte verwenden Funkstandards wie NarrowBand-IoT. Alle Komponenten senden ihre Daten an ein Gateway, das mit der Middleware der Smart City-Plattform verbunden ist. Die Middleware ermöglicht die Kommunikation der IoT-Komponenten untereinander und mit den Plattform-Diensten. Um die gesendeten Datenmengen zu reduzieren, können erste Teile der Datenanalyse bereits auf dem Gateway realisiert werden. Hierfür muss vorab entschieden werden, welche Daten wichtig genug sind, dass sie über die SIM-Karte bzw. die Netzwerkverbindung an die Big Data-Plattform gesendet werden. Die dort aufbereiteten Daten werden von der Solution Enabling-Plattform und der darauf befindlichen Business Logik weiter verarbeitet und visualisiert. Der modulare Aufbau der Plattform kann mit Hilfe von Integrationslösungen,

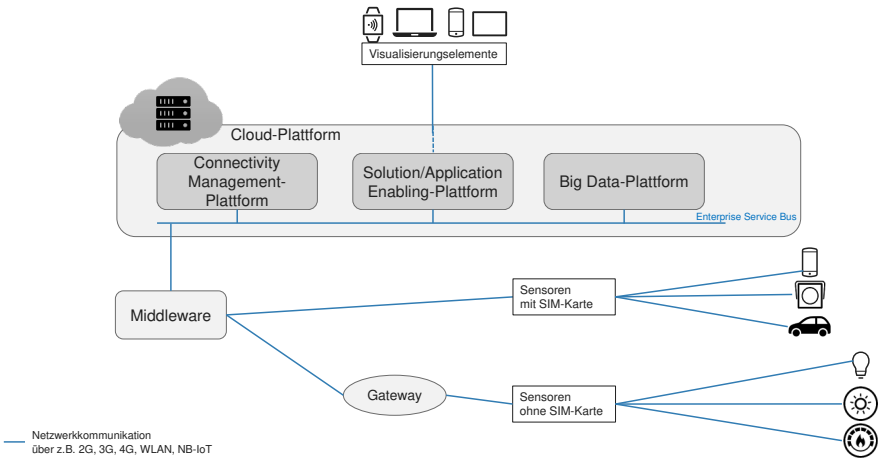


Abb. 1: Typische Smart City-Plattform

wie z.B. dem Enterprise Service Bus (ESB) oder durch Microservices und serverless-Architekturen, ermöglicht werden. Alternativen zum ESB sind die Lightweight Internet of Things Service Bus Architecture [Ne15] oder MuleESB [Br09], welche sich speziell für den Einsatz im IoT eignen. Ein Mediation-Service dient als Vermittler, welcher die einzelnen Plattformen – unabhängig vom Standort – miteinander verbindet. Über so eine Integrationslösung wird auch die Kommunikation zwischen den Plattformen ermöglicht.

2.3 Ende-zu-Ende-Sicherheit im Internet of Things

Da die IoT-Geräte und die Smart City-Plattform oft von verschiedenen Akteuren betrieben werden, schlägt [Ci18] differenzierte Eigentums- und Richtlinienkonzepte mit Zugriffskontrollen an den Schnittstellen vor. Damit die beteiligten Akteure Daten austauschen können, müssen Vertrauensbeziehungen zwischen ihnen aufgebaut werden. Da IoT-Komponenten nur über eingeschränkte Ressourcen verfügen, können klassische Authentifizierungs- und Verschlüsselungsverfahren wie AES oder RSA nicht eingesetzt werden. Daher schlägt [Ci18] leichtgewichtige Authentifizierungsverfahren vor. [OC18] definiert Sicherheit im IoT auf Basis unterschiedlicher Verantwortlichkeiten der Beteiligten im IoT-Ökosystem. Hardwarehersteller, Applikationsentwickler, Verbraucher, Betreiber und weitere Beteiligte sind dafür verantwortlich, dass Prozesse zur Erreichung der Sicherheit im IoT umgesetzt werden. Angriffe auf IoT-Geräte können die unbefugte Beschaffung von sensiblen oder privaten Daten, deren Manipulation sowie das Stören oder Verhindern von Services innerhalb des IoT-Systems mit sich bringen. Die von [OC18] vorgeschlagenen Lösungen sind jedoch nur auf einzelne Bestandteile des IoT-Ökosystems beschränkt und bilden kein E2E-Sicherheitskonzept ab.

[BA11] skizziert ein IoT-Sicherheitsframework auf Basis von leichtgewichtiger Kryptografie, physikalischer Sicherheit über ein Trusted Plattform Modul, standardisierten Sicherheitsprotokollen, sicheren Betriebssystemen, dem berücksichtigen von zukünftigen Anwendungsbereichen und sicheren Speichern. Allerdings bleibt offen, wie diese Merkmale implementiert werden können. Die Arbeit beschränkt sich auf generische Sicherheits-Ansätze für die im IoT angewendeten Protokolle, sowie für die Hard- und Softwareplattformen.

[BM16] beschreibt ein Sicherheitsframework auf verschiedenen Layern, welches auf der Blockchain-Technologie basiert. Die Layer präsentieren eine Prozesskette von den Sensoren und Aktoren über einen Kommunikations-Layer bis zu einem Layer für die Applikationen. Die Transaktionsdaten werden in einer Blockchain im Database-Layer gespeichert. Es wird jedoch nicht berücksichtigt, dass der Energiebedarf und der Overhead einer Blockchain eine Implementierung im IoT nicht unmittelbar zulassen. [Do17] definiert eine leichtgewichtige Variante der Blockchain für die Absicherung von IoT-Transaktionen. Eine Evaluierung anhand eines Smart Home-Szenarios zeigt, dass die angestrebten Schutzziele auch erreicht werden können. Es bleibt jedoch offen, ob eine Implementierung der Blockchain direkt innerhalb einer Smart Home-Anwendung möglich ist und ob sich diese Blockchain-Variante für weitere IoT-Anwendungsfälle eignet.

3 Die multimodale Mobilität

Der Nutzer legt einmalig über eine App auf seinem Smartphone oder über eine Webseite mit einer Webanwendung ein Benutzerprofil mit Zahlungsinformationen und Login-Daten an. Ist dies geschehen, läuft eine Reise wie folgt ab:

- Der Nutzer kann über eine App oder Webanwendung Routen mit verschiedenen Reisemitteln planen. Dafür werden Positions- und Verfügbarkeitsdaten benötigt.
- Tickets und Reservierungen werden in der Smart City-Plattform gebucht und über einen Abrechnungsdienstleister bezahlt. Hierbei werden die Zahlungsinformationen des Nutzers benötigt.
- Das Benutzerprofil stellt der Plattform alle Daten für den Buchungsprozess bereit.
- Buchungsinformationen inkl. Zahlungsinformationen und Nutzungsdaten werden an den jeweiligen Mobilitätsanbieter weitergeleitet.
- Positions-, Nutzungs- und Verfügbarkeitsdaten der Verkehrsmittel werden von IoT-Komponenten gesammelt und in einer Cloud weiterverarbeitet.
- Das Vernetzen der Verkehrsmittel ermöglicht eine Positionsberechnung in Echtzeit, um Verspätungen einzuplanen oder Wegezeiten zu aktualisieren.
- Sobald Anschlüsse gefährdet sind, werden auf der Basis von Positionsdaten und Daten über die Reiseroute Alternativrouten ermittelt.
- Mobilitätsanbieter nutzen die Daten, um Echtzeit-Fahrpläne zu aktualisieren.
- Die Stadtplanung nutzen die Daten, um das Mobilitätsangebot zu optimieren.

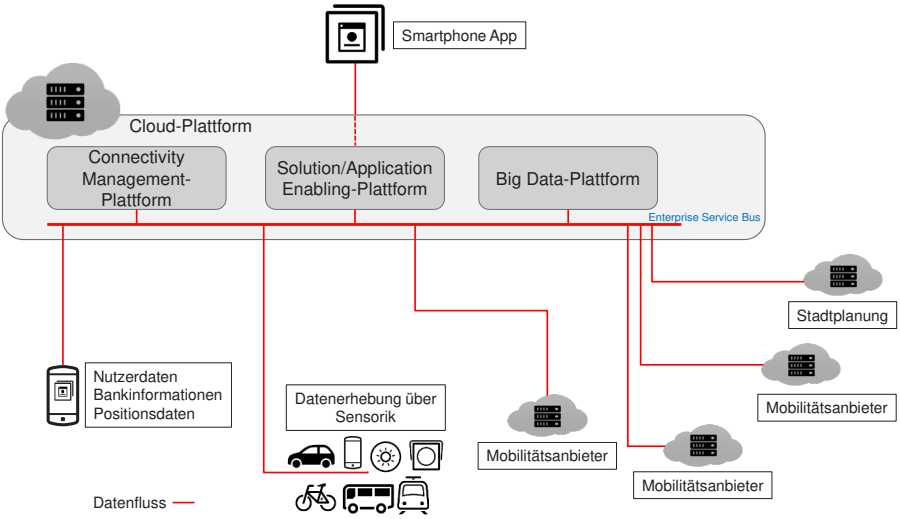


Abb. 2: Datenflüsse der multimodalen Mobilität

3.1 Datenflüsse und Wertschöpfungskette

Abbildung 2 illustriert die Datenflüsse bei der multimodalen Mobilität. Die IoT-Hardware kann in Fahrrädern, Autos, Bussen oder in Straßenbahnen verbaut sein. Auch das Mobiltelefon eines Nutzers kann als IoT-Komponente den Standort des Nutzers ermitteln. Jede Komponente nutzt verschiedene Kommunikationskanäle, um Daten zur Erfüllung der eigenen Funktion auszutauschen. Dabei werden die Mobilfunkstandards 2G, 3G und 4G sowie WLAN und kabelgebundene Übertragungswege genutzt. Abbildung 3 zeigt die Wertschöpfungskette der multimodalen Mobilität in Form einer Matrix, die den Komponenten, auf denen eine Wertschöpfung stattfindet, Rollen, Kommunikationskanäle und Daten zuordnet. Die Übersicht wurde auf Basis von Abschnitt 2 erstellt.

Welcher Schutzbedarf für diese Daten und alle Komponenten und Anwendungen besteht, die auf sie zugreifen, wird mit einer Schutzbedarfsfeststellung nach BSI-Standard 200-2 [Bu17a] ermittelt. Dazu ist eine Strukturanalyse erforderlich, welche die Daten, Dienste, Übertragungswege etc. modelliert. Aus Platzgründen verzichten wir auf eine detailliert Darstellung der Strukturanalyse sowie der Schutzbedarfsfeststellung und verweisen auf unseren Bericht [PB19]. Wir haben 6 Kategorien von Daten ermittelt, nämlich *Verfügbarkeitsdaten* (D1) der Verkehrsmittel, *Positionsdaten der IoT-Geräte* (D2), *Zahlungsdaten* (D3), *Login-Daten* (D4), *Positionsdaten der Nutzer* (D5) und *Nutzungsdaten* (D6) zur Abrechnung. Diese Daten werden von Webanwendungen, Smartphone Apps, Datenbanken und Big-Data-Analyseverfahren auf IoT-Geräten, Smartphones und der Smart City-Plattform verarbeitet und über WLAN, Mobilfunk, rechenzentrumsinterne Verbindungen, das Internet und Enterprise Service Bus-Verbindungen übertragen.

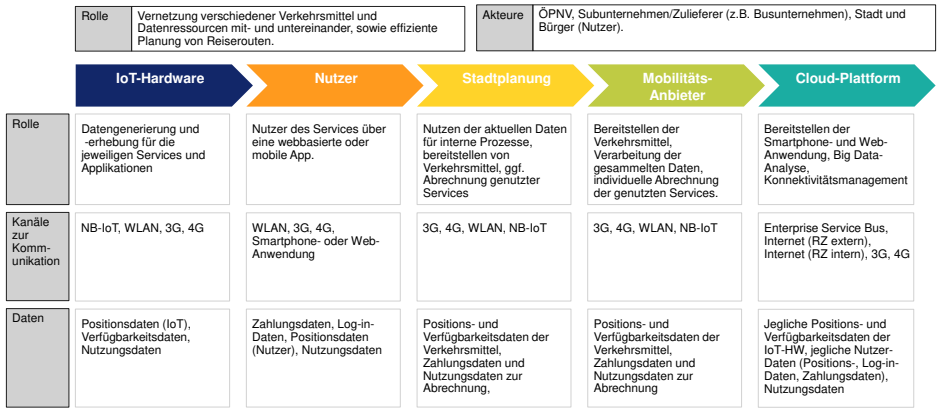


Abb. 3: Wertschöpfungskette der multimodalen Mobilität

Die Schutzbedarfsfeststellung unterscheidet *Vertraulichkeit*, *Integrität* und *Verfügbarkeit* der zu schützenden Objekte. [Bu17a] definiert dabei die Schutzbedarfe „normal“, „hoch“ und „sehr hoch“. Der Schutzbedarf für eine IT-Komponente wird anhand der verwendeten Daten bestimmt. Da unsere Daten zur Planung und Abrechnung verwendet werden, haben wir keine existenzbedrohenden Schadensszenarien oder solche, die zu schweren gesundheitlichen Schäden führen, festgestellt. Wir haben den Schutzbedarf „sehr hoch“ also nicht vergeben. Dagegen können durch fehlerhafte Planung oder Abrechnung erhebliche materielle Schäden auftreten. Daher haben wir vielen Daten und davon abgeleitet auch vielen Komponenten den Schutzbedarf „hoch“ in den Zielen Verfügbarkeit, Vertraulichkeit und Integrität zugewiesen. Den Schutzbedarf „normal“ haben von uns die Verfügbarkeits- und Positionsdaten (D1, D2) in der Kategorie Vertraulichkeit erhalten, weil hier keine personenbezogenen Daten verarbeitet werden. Des weiteren haben wir den Zahlungs-, Login- und Positionsdaten der Nutzer (D3, D4, D5) bei der Integrität den Schutzbedarf „normal“ zugewiesen, weil Fehler in diesen Daten aufgrund Eigenschaften der Sensorik ohnehin berücksichtigt werden müssen, nur geringe Auswirkungen haben, oder leicht entdeckt und behoben werden können.

Nutzt eine IT-Anwendung Daten mit unterschiedlichen Schutzbedarfen, erhält sie den höchsten dieser Schutzbedarfe. Ebenso wird der Schutzbedarf der Kommunikationswege, IT-Systeme und Plattformen bestimmt. Daher erhalten alle in der Cloud-Plattform ablaufenden Anwendungen sowie deren Kommunikationswege den Schutzbedarf „hoch“ für Verfügbarkeit, Vertraulichkeit und Integrität. Aufgrund des Schutzbedarfs „hoch“ ist eine Risikoanalyse [Bu17b] erforderlich. Für weiterführende Details sowie für die Kritikalität der Kommunikationsverbindungen und eine Zuordnung von Elementargefährdungen wie „Feuer“ oder „Offenlegung schützenswerter Informationen“ verweisen wir auf unseren Bericht [PB19].

3.2 Risikoanalyse

Die Risikoanalyse hat das Ziel, alle anwendungsfallspezifischen Risiken aufzudecken, die von den im Grundschutz-Kompendium [Bu19] enthaltenen Standard-Maßnahmen nicht abgedeckt werden. Die Risikoanalyse wurde gemäß [Bu17b] in Form von Interviews im Workshop-Charakter mit zwei Partnern der Detecon International GmbH durchgeführt, die Experten der Themen IoT, Smart City und Connected Car sind und aufgrund langjähriger Berufserfahrung über ein breites Wissen im Bereich Netzwerk- und Plattformsicherheit verfügen. Wir haben die in Tabelle 1 aufgelisteten zusätzliche Gefährdungen für den gesamten Informationsverbund ermittelt. Für die IoT-Komponenten haben wir die in Tabelle 2 aufgeführten zusätzliche Gefährdungen identifiziert.

Gesamter Informationsverbund	
Name	Beschreibung
G z.1: Abfangen der Daten entlang der Wertschöpfungskette	Bei mangelnden Sicherheitsmaßnahmen lässt sich der Übertragungsweg der Daten D3, D4, D5 und D6 unbemerkt unterbrechen. Ein Dienstleister könnte z.B. nicht zwischen „keine Zahlung“ und „Zahlungsdaten abgefangen“ unterscheiden.
G z.2: Abhören/Lesen der Daten entlang der Wertschöpfungskette	Bei mangelnden Sicherheitsmaßnahmen lassen sich die Daten auf den IoT-Komponenten in der Wertschöpfungskette unbemerkt mitlesen, z.B. zu präferierten Routen (D5 und D6) oder Zahlungsinformationen (D3, D4 und D6) des Nutzers.
G z.3: Manipulieren der Daten innerhalb der Wertschöpfungskette	Bei mangelnden Sicherheitsmaßnahmen lassen sich die Daten von IoT-Komponenten entlang der Wertschöpfungskette unbemerkt manipulieren.

Tab. 1: Zusätzliche Gefährdungen im Informationsverbund

IoT-Komponenten	
Name	Beschreibung
G z.4: Unbefugte Übernahme der IoT-Komponente	Kann ein Angreifer eine IoT-Komponente übernehmen, kann er das dahinterliegende IT-System beeinflussen.
G z.5: Manipulation der IoT-Komponente	Kann ein Angreifer mit physischem Zugang oder über Zugang zum Kommunikationskanal eine IoT-Komponente manipulieren, ist deren Funktion nicht mehr gewährleistet.
G z.6: Sichtbarkeit der IoT-Komponente nach außen/extern	Sollte eine IoT-Komponente eine aus dem Internet sichtbare IP-Adresse besitzen, kann sie möglicherweise auch über das Internet angegriffen werden.

Tab. 2: Zusätzliche Gefährdungen für IoT-Komponenten

Für alle zusätzlichen Gefährdungen finden sich im BSI Grundschutz-Kompendium entsprechende Elementargefährdungen. Allerdings geht das BSI für unser Szenario nicht auf den Sonderfall „IoT-Komponente“ ein. Um selbst angemessene Maßnahmen festzulegen, muss das Risiko ermittelt werden, das von den Gefährdungen ausgeht. Dabei wird zwischen der Eintrittshäufigkeit und der potentiellen Schadenshöhe unterschieden.

Wir haben die Differenzierung in die Schadenshöhen „normal“, „hoch“ und „sehr hoch“ aus [Bu17b] übernommen. Bei den Eintrittshäufigkeiten unterscheiden wir zwischen „begrenzt“,

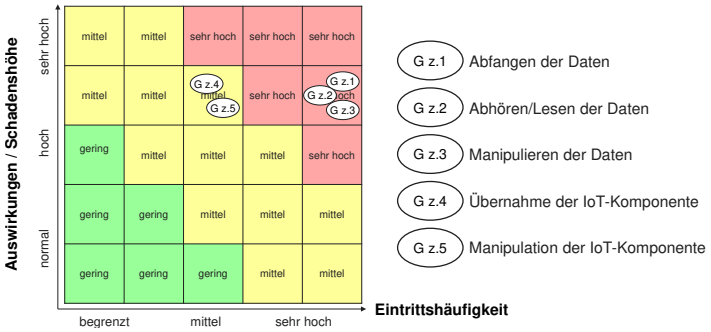


Abb. 4: Übersicht über die Risiken

„mittel“ und „sehr hoch“. Mit diesem Maßstab haben wir für die Gefährdungen G z.1 „Abfangen von Daten“, G z.2 „Abhören/Lesen der Daten“ und G z.3 „Manipulieren der Daten“ die Eintrittshäufigkeit „sehr hoch“ und die Schadenshöhe „hoch“ festgelegt, da Komponenten teilweise aus dem Internet sichtbar sind und personenbezogene Daten verarbeiten. Für G z.4 „Unbefugte Übernahme der IoT-Komponente“ und G z.5 „Manipulation IoT-Komponente“ haben wir die Eintrittshäufigkeit „mittel“ vergeben, da der Angreifer physischen Zugriff benötigt oder von innen kommen muss. Die potentielle Schadenshöhe haben wir mit „hoch“ festgelegt, da ein Angreifer auf diese Weise erhebliche finanzielle Schäden, Image-Schäden und Rechtsverletzungen verursachen kann. Abbildung 4 gibt einen Überblick über diese Risiken. „G z.6 Sichtbarkeit der IoT-Komponente“ haben wir nicht in Abbildung 4 aufgeführt. G z.6 erhöht die Eintrittswahrscheinlichkeit für viele Elementargefährdungen und muss durch Maßnahmen abgesichert werden, ist jedoch auf einer allgemeinen Ebene nicht mit einer potentiellen Schadenshöhe zu bewerten. Details und Begründungen sind wieder in [PB19] nachzuschlagen.

4 Risikobehandlung

Aus der Wertschöpfungskette der multimodalen Mobilität und den damit einhergehenden Datenflüssen ergeben sich fünf E2E-Beziehungen, die abgesichert werden müssen:

1. zwischen Nutzer und Mobilitätsanbieter
2. zwischen Nutzer und Abrechnungsdienstleister
3. zwischen IoT-Komponente und Smart City-Plattform
4. zwischen Stadtplanung und Smart City-Plattform
5. zwischen Mobilitätsanbieter und Smart City-Plattform

Dabei schützt der APN den Übertragungsweg, indem er eine Trennung zwischen den privaten Adressen der IoT-Komponenten (mit SIM-Karte) und öffentlich sichtbaren IP-Adressen schafft. Auch innerhalb der Internetverbindung von externen Plattformen zur Smart City-Plattform existieren Datensicherheitsmaßnahmen wie beispielsweise VPNs

oder X.509-authentifizierte Verbindungen. Allerdings berücksichtigen diese Maßnahmen die E2E-Sicherheit nicht. Bezieht man die im letzten Abschnitt identifizierten Risiken auf unsere fünf E2E-Beziehungen, lassen sich folgende Handlungsbedarfe identifizieren:

IoT-Komponente: Die IoT-Komponente ist dem Angreifer zugänglich, wodurch physische Manipulationen oder Sabotage möglich werden. Teilweise sind IoT-Komponenten aus dem Internet erreichbar, wodurch auch Manipulationen ohne physische Nähe zur Komponente möglich werden. Der Anwendungsfall benötigt jedoch genaue Daten, um Routen zu planen oder Abrechnungen durchzuführen.

Mobilfunk: Die Daten sind zwar innerhalb des Mobilfunknetzes mit einem Mobilfunk-Verschlüsselungsverfahren verschlüsselt. Allerdings sind die Verschlüsselungsverfahren des GSM- und UMTS-Netzes bereits gebrochen, weswegen lediglich der aktuelle LTE-Verschlüsselungsalgorithmus einen Schutz gegen das unerlaubte Lesen und Abhören von Daten bietet. Bei nicht verfügbarem LTE wird auf das unsichere GSM oder UMTS zurückgegriffen. Eine E2E-Datensicherheit ist nicht gewährleistet.

Smart City-Plattform: Die Absicherung der Übertragungswege schützt die ausgetauschten Daten nur bis zur Schnittstelle der Plattform. Es existiert keine durchgängige Verschlüsselung der Daten beispielsweise zu einer verschlüsselten Datenbank. So ist nicht sichergestellt, dass Daten nicht verändert oder von einem nicht autorisierten Gerät gesendet wurden.

4.1 Übertragungssicherheit

Eine Option zur Risikobehandlung bietet das Internet Protokoll Version 6 (IPv6). IPv6 baut ein Mesh-Netzwerk auf, das Daten mit IPsec-Verschlüsselung überträgt. Dabei können private IP-Adressen verwendet werden, wodurch das Gerät aus dem Internet nicht sichtbar ist. Allerdings ist IPv6 nicht mit IPv4 kompatibel und nicht überall verfügbar. Es wird daher eine Maßnahme gesucht, die die Datensicherheit in existierenden (IoT-)Netzwerken realisieren kann. Nachfolgend diskutieren wir, ob die Verwendung der Transport Layer Security (TLS)-Verschlüsselung für eine E2E-Absicherung geeignet ist.

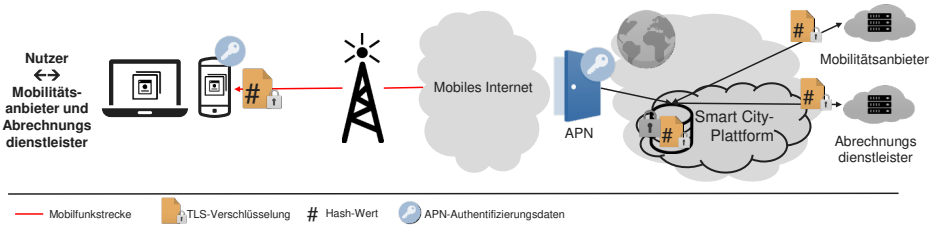


Abb. 5: E2E-Absicherung zum Nutzer

Die Abbildungen 5 bis 7 zeigen unsere E2E-Beziehungen mit integrierter TLS-Verschlüsselung. Die TLS-Verschlüsselung muss jeweils in den Endpunkten der Beziehungen

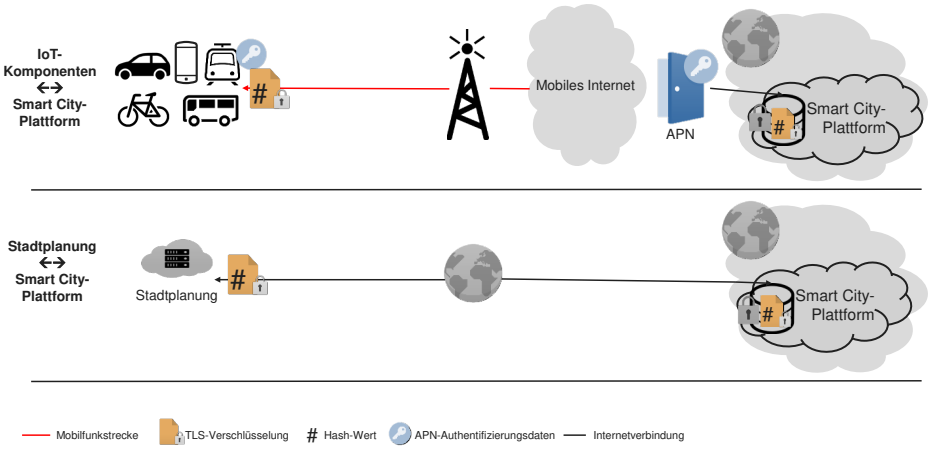


Abb. 6: E2E-Absicherung zu IoT-Komponenten und Stadtplanung

implementiert sein. In Abbildung 5 sind dies die Web-Anwendung bzw. Smartphone-App sowie die Smart City-Plattform und die Dienstleister. Werden Daten von den Anwendungen bzw. vom Nutzer generiert, werden diese mittels TLS verschlüsselt. Die verschlüsselten Daten werden über das Mobilfunknetz zur Basisstation gesendet. Das IoT-Gerät authentifiziert sich dann am APN mit dem jeweiligen Anmeldenamen. Die Smart City-Plattform entschlüsselt die TLS-verschlüsselten Daten, speichert sie in einer Datenbank und stellt sie anderen Services innerhalb der Plattform zur weiteren Verarbeitung zur Verfügung. Sollten Daten aus der Datenbank an externe Plattformen, hinsichtlich Abrechnung oder Reservierungen, gesendet werden, geschieht dies ebenfalls durch eine TLS-Verschlüsselung.

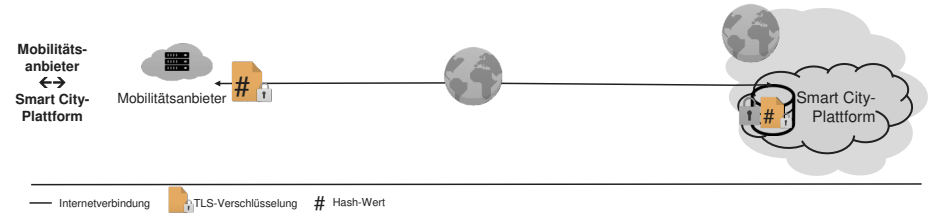


Abb. 7: E2E-Absicherung zum Mobilitätsanbieter

Die Absicherung der anderen E2E-Beziehungen erfolgt analog. Durch diese Maßnahme ist auch nach einem Bruch der LTE-Verschlüsselung oder bei einem Man-in-the-Middle-Angriff auf dem Gateway die Übertragungssicherheit sichergestellt.

4.2 Manipulationssicherheit

Da bei der multimodalen Mobilität zahlreiche Akteure miteinander interagieren, muss auch sichergestellt werden, dass die übermittelten Daten protokoll- und spezifikationsgerecht verarbeitet werden (G z.3 „Manipulieren der Daten“). Klassisch lässt sich dies über eine Trusted Third Party lösen, die Prüfsummen von allen erzeugten Daten speichert. Das heißt, bei jeder Datenübermittlung wird zusätzlich ein kryptographischer Hash-Wert als Prüfsumme über die Daten berechnet und ebenfalls TLS-verschlüsselt an eine Datenbank innerhalb der Smart City-Plattform gesendet, die hier als Trusted Third Party auftritt (s. Abbildungen 5 bis 7). Die Smart City-Plattform muss sicherstellen, dass die Datenbank nicht nachträglich verändert werden kann und jeder Akteur Zugang zu den Hash-Werten hat. In einer offenen Smart City-Umgebung ist es jedoch möglicherweise nicht wünschenswert, einen einzelnen Akteur als Trusted Third Party zu etablieren, der zugleich ein Plattformbetreiber ist und damit einen kritischen Angriffspunkt darstellt.

Eine Alternative könnte das Blockchain-Protokoll bieten. Ein Blockchain-Block [BM16] enthält neben den Transaktionsdetails Informationen über die Blocknummer, Prüfsummen über den Vorgängerblock sowie Informationen zur Validierung dieses Blocks. In einem E2E-Sicherheitskonzept könnte eine Blockchain analog zu [Fl18] Manipulationssicherheit herstellen: Die Daten werden, nachdem sie durch das IoT-Gerät generiert wurden, wie beschrieben mit TLS verschlüsselt. Zusätzlich wird mit einer kryptografischen Hash-Funktion eine Prüfsumme über die Daten berechnet. Die Prüfsumme wird dann nicht in einer Datenbank auf der Plattform, sondern mit einem Zeitstempel versehen in einer Blockchain gespeichert. Dies ermöglicht allen Akteuren jederzeit eine Verifizierung der erhaltenen Daten, ohne dabei auf zentralisierte Verfahren zurückgreifen zu müssen.

4.3 Diskussion

Die Gefährdungen G z.1, G z.2 und G z.3 werden mit den von uns betrachteten Maßnahmen erheblich reduziert. Die TLS-Verschlüsselung verhindert das Lesen, Abhören und Abfangen der Daten, eine kryptographische Prüfsumme deren Manipulation. Die Gefährdungen G z.4 und G z.5 können nicht mit den vorhandenen technischen Maßnahmen abgesichert werden. Allerdings kann hier mit manuellen und prozessualen Sicherheitsmaßnahmen entgegengewirkt werden. Hierzu zählen security-by-design-Entwicklungsansätze für die Soft- und Hardware der IoT-Komponenten.

Ein Grundschutz lässt sich also erreichen, indem alle institutionellen Teilnehmer dazu verpflichtet werden, passende ISO-Standards und ausreichende Zertifizierungen nachzuweisen. Für die Nutzer gilt dies jedoch nicht: Sie verwenden eigene Geräte, um die angebotenen Dienste zu nutzen, und müssen daher selbst für die Absicherung ihres Endpunkts sorgen. Dadurch entstehende Sicherheitsprobleme sind jedoch auf den individuellen Nutzer mit seinem unzureichend abgesicherten Gerät beschränkt. Der Schaden bei einem Sicherheitsvorfall ist also auf einzelne Nutzer begrenzt. Dies ließe sich durch vertragliche Maßnahmen auffangen, vergleichbar zu den Stornierungsmöglichkeiten bei unberechtigten Kreditkartenbuchungen.

5 Fazit

Die Digitalisierung alltäglicher Prozesse führt zu Herausforderungen für Datenschutz und Datensicherheit. Die Nutzer solcher Prozesse sind darauf angewiesen, dass die institutionellen Akteure mit diesen Herausforderungen geeignet umgehen. Die Analyse des Anwendungsfalls „Multimodale Mobilität“ nach den BSI-Standards 200-2 und 200-3 hat eine Reihe von Gefährdungen und Handlungsbedarfe aufgezeigt. Zwar enthält das IT-Grundschutz-Kompendium Bausteine zum Erreichen der Datensicherheit, jedoch machen die Besonderheiten des IoT-Einsatzes in einer Smart City detaillierte Risikoanalysen notwendig. Dabei ist das Ineinandergreifen von Sicherheitsmaßnahmen zwischen verschiedenen Akteuren über Unternehmensgrenzen hinweg eine Herausforderung. Sollte ein beteiligtes Unternehmen Sicherheitsmaßnahmen unvollständig implementieren, kann keine umfassende Datensicherheit für alle Nutzer garantiert werden. Wir konnten zeigen, dass sich ein hohes Maß an IT-Sicherheit bei der multimodalen Mobilität realisieren lässt, wenn alle institutionellen Akteure an der Wertschöpfungskette auf die Umsetzung des aktuellen Stands der Technik verpflichtet werden. Dies begrenzt Schadensfälle auf die privaten Mobilgeräte einzelner Nutzer, die selbst über die Sicherheitsmerkmale ihres Geräts verfügen können.

Literatur

[BA11] BABAR, S. et al.: Proposed Embedded Security Framework for Internet of Things. In: Conference on Wireless Communication, Vehicular Technology, Information Theory and Aerospace & Electronic Systems Technology. 2011.

[BM16] Biswas, K.; Muthukkumarasamy, V.: Securing smart cities using blockchain technology. In: 2016 IEEE 18th international conference on high performance computing and communications; IEEE 14th international conference on smart city; IEEE 2nd international conference on data science and systems (HPCC/SmartCity/DSS). IEEE, S. 1392–1393, 2016.

[Bo19] Bourne, V.: Unternehmen vernachlässigen IoT-Sicherheit und setzen das Vertrauen der Kunden aufs Spiel, https://www.trendmicro.com, Kopie s. http://www.webcitation.org/75F0WF9y4, 2019.

[Br09] Brebner, P.: Service-oriented performance modeling the mule enterprise service bus (esb) loan broker application. In: Euromicro Conference on Software Engineering and Advanced Applications. 2009.

[Bu17a] Bundesamt für Sicherheit in der Informationstechnik: BSI-Standard 200-2, IT-Grundschutz-Methodik. https://www.bsi.bund.de/, 2017.

[Bu17b] Bundesamt für Sicherheit in der Informationstechnik: BSI-Standard 200-3, Risikomanagement. https://www.bsi.bund.de/, 2017.

[Bu18] Bundesamt für Sicherheit in der Informationstechnik: Die Lage der IT-Sicherheit in Deutschland 2018. https://www.bsi.bund.de/, 2018.

[Bu19] Bundesamt für Sicherheit in der Informationstechnik: IT-Grundschutz-Kompendium - Edition 2019. https://www.bsi.bund.de/, 2019.

[Ci18] Cisco: Securing the Internet of Things: A Proposed Framework, https://www.cisco.com, Kopie s. http://ww.webcitation.org/72ffITErR, 2018.

[Do17] Dorri, A.; Kanhere, S. S.; Jurdak, R.; Gauravaram, P.: Blockchain for IoT security and privacy: The case study of a smart home. In: IEEE Pervasive Computing and Communications Workshops. 2017.

[Ec18] Eckert, C.: IT-Sicherheit: Konzepte-Verfahren-Protokolle. Walter de Gruyter, 2018.

[Fl18] Florea, B. C.: Blockchain and Internet of Things data provider for smart applications. In: 2018 7th Mediterranean Conference on Embedded Computing (MECO). IEEE, S. 1–4, 2018.

[GB14] Gallotti, R.; Barthelemy, M.: Anatomy and efficiency of urban multimodal mobility. Nature Scientific Reports 4/, S. 6911, 2014.

[HL17] Haritha, A.; Lavanya, A.: Internet of Things: Security Issues. International Journal of Engineering Science Invention 6/11, 2017.

[JS14] Jochema, P.; Schipplb, J.: 8. Mobility 2.0: Antriebskonzepte im Zusammenspiel mit multimodaler Mobilität. ALTERNATIVE/, S. 165, 2014.

[Ma06] Maertins, C.: Die Intermodalen Dienste der Bahn: mehr Mobilität und weniger Verkehr? Wirkungen und Potenziale neuer Verkehrsdienstleistungen. Discussion Papers / Wissenschaftszentrum Berlin für Sozialforschung gGmbH. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-113845, 2006.

[Ne15] Negash, B.; Rahmani, A.-M.; Westerlund, T.; Liljeberg, P.; Tenhunen, H.: LISA: Lightweight internet of things service bus architecture. Procedia Computer Science 52/, S. 436–443, 2015.

[OC18] O'Connor, C.: Security in the era of cognitive IoT, https://www.ibm.com/blogs, Kopie s. http://www.webcitation.org/6z2eymUj9, 2018.

[PB19] Plate, F.; Buchmann, E.: Ende-zu-Ende-Sicherheit für die Multimodale Mobilität in einer Smart City, Technischer Bericht, http://nbn-resolving.de/urn:nbn:de:bsz:14-qucosa2-337880, Hochschule für Telekommunikation Leipzig, 2019.

[PP+16] Patel, K. K.; Patel, S. M. et al.: Internet of things-IOT: definition, characteristics, architecture, enabling technologies, application & future challenges. International journal of engineering science and computing 6/5, 2016.

[Xi12] Xia, F.; Yang, L. T.; Wang, L.; Vinel, A.: Internet of things. International Journal of Communication Systems 25/9, S. 1101, 2012.

Track 6 - Digitalisierung des Energiesystems

Digitization of the Energy System

Kurt Rohrig[1], Christoph Krauß[2]

Europe's energy supply is characterized by the comprehensive transformation of structure, processes and business models. The unbundling and the merging of networks and markets have shaped this development as well as the enormous growth of renewable energies and the associated decentralization. This transformation process can only succeed if the massive use of modern information and communication technology (ICT) and its innovations are used in such a way that the pillars of the energy triangle are pursued and implemented in a balanced manner. The fast-moving digitization and automation process will be the companion to transforming our energy system, which will be highly flexible and decentralized. The massive interaction between generation, transport, storage and consumption requires digitization by intelligent components as well as networking of the overall system in the superordinate level. The process involves developing disruptive technologies as well as software architectures, applications and new functionalities of intelligent components.

This track gives an overview of the challenges and opportunities as well as the current achievements of using modern digital solutions in the field of energy supply. Particular attention is paid to the integration of renewable energies and the transformation of the entire energy supply system. Current developments for improved wind and solar power forecasts, optimized grid operation, efficient use of balancing power, new market mechanisms and solutions for low-emission electricity use are presented. A keynote introduces the new meaning of mass data and data science.

The Program Committee consists of

Andreas Fuchs, Fraunhofer SIT

Reinhard Mackensen, Fraunhofer IEE

Roland Rieke, Fraunhofer SIT

Mathias Uslar, OFFIS

[1] Fraunhofer IEE, Kassel, kurt.rohrig@iee.fraunhofer.de
[2] Fraunhofer SIT, Darmstadt, Christoph.krauss@sit.fraunhofer.de

doi:10.18420/inf2019_67

Invited Presentations

Is Data Oxygen?

Marc Peters[1]

Sind sie gelangweilt von all dem Hype rund um Machine Learning und Künstliche Intelligenz und dem Heilsversprechen für die Energie- und Umweltindustrie?

Beginnend mit Beispielen wie und wo Daten einen maßgeblichen Einfluss schon heute auf uns Menschen haben in Projekten im Bereich Energie und Umwelt begeben sie sich auf eine Reise zu den 6-Ds um die Ausgangsfrage der Session zu klären - ‚**Is Data Oxygen?**'.. Mit Projektbeispielen und Erfahrungen wird dabei das Thema praktisch greifbar und Technologie verständlich dargestellt.

Die Session wird abgerundet mit einem Einblick in einige ‚Summer of `69 ' Ereignisse und Innovationen welche einen Bezug zu den heutigen Herausforderungen der E,E&U Industry haben, diese beeinflussen oder unterstützen.

1 IBM Deutschland GmbH, Gustav-Heinemann-Ufer 120-122 D-50968 Köln. Marc.Peters@de.ibm.com

doi:10.18420/inf2019_68

Full Papers

Development and Application of KPIs for the Evaluation of the Control Reserve Supply by a Cross-border Renewable Virtual Power Plant

Julia Strahlhoff[1], Andreas Liebelt[1], Stefan Siegl[1] and Simon Camal[2]

Abstract: In an increasingly decentralised energy system with a rising share of fluctuating renewable energy sources, such as wind energy and photovoltaics (PV), the importance of virtual power plants (VPP) for the provision of ancillary services is growing. Transmission system operators (TSOs) impose stringent requirements on the reliability and accurate controllability of power plants that are susceptible to qualify for control reserve provision. It needs to be discussed whether these requirements have to evolve to enable the participation of renewables in control reserve markets. As part of the REstable research project, an innovative ICT infrastructure was set up for linking German and French wind and PV farms in a transnational VPP. Key performance indicators (KPIs) derived from technical requirements of the German TSOs are proposed and applied to quantify the quality of control reserve provision by the VPP in physical field tests.

Keywords: Virtual Power Plants; System Architecture; Control Reserve; Requirements; KPIs

1 Introduction

In the entire synchronous area of the ENTSO-E (European Network of Transmission System Operators for Electricity), the grid frequency must remain within precisely defined limits at all times in order to avoid consumer disconnections and grid breakdowns [Sw06]. For this reason control reserve is used to keep the grid frequency stable at 50 Hertz. The TSOs have the responsibility to organise markets for the provision of positive and negative control reserve in three different qualities (Frequency Containment Reserve, automatic and manual Frequency Restoration Reserve) in their control zones and to apply them at short notice if necessary [Eu17]. The fluctuations and forecasting errors in wind and PV generation will increase the need for flexibility and reserves in the European electricity grid with the rising share of renewable energies [Ac15]. The aggregation of renewable power plants can reduce their production uncertainty and enables aggregated plants controlled via a VPP to submit control reserve bids based on e.g. probabilistic forecasting [CMK18]. The general capabilities of renewable energies as well as control mechanisms and technical challenges for providing ancillary services have already been reviewed and discussed e.g. in [ADA18], [HSS16] and [Dí14]. Moreover, the

[1] Fraunhofer Institute for Energy Economics and Energy System Technology (IEE), Königstor 59, 34119 Kassel, Germany, {julia.strahlhoff; andreas.liebelt ; stefan.siegl}@iee.fraunhofer.de
[2] MINES ParisTech – PSL University, Rue Claude Daunesse, CS 10207, 06904 Sophia Antipolis Cedex, France, simon.camal@mines-paristech.fr

doi:10.18420/inf2019_69

physical ability of wind energy and photovoltaics to provide control reserve within a VPP has already been demonstrated in former research projects of Fraunhofer IEE [Fr14], [Fr17]. In these projects the control of energy units was realized via the in-house software for a central control system «IEE.vpp».

A new challenge is to build up the ICT connection for a various set of power plants from different manufactures and operators in different countries and to overcome the country specific regulatory obstacles for a central control of wind and PV farms. Last, it is a challenge to prove the fulfilment of the strictest requirements for the provision of Frequency Containment Reserve (FCR) with fluctuating renewables. The German-French-Portuguese project «REstable» takes up these challenges to improve renewable-based system services by better cooperation of the European control zones.

The objective of this work is to analyse the quality of control reserve provided by fluctuating energies in an exactly quantifying and comparable way and to compare the results for the currently valid technical requirements with the newly published prequalification conditions by the German TSOs [Ge18b]. Physical control power field tests carried out with IEE.vpp are evaluated with the help of a key figure system that is developed within the context of the REstable project. The mathematical formulation of the KPIs is derived from the requirements in the publicly accessible documents of the German TSOs. With the use of Python automated evaluations are visualized.

Since FCR operates across countries in the entire network system and places the greatest requirements on control speed and technology, the focus of this work is on the execution and evaluation of FCR field tests. The FCR field tests in the REstable project explore possible pre-qualification frameworks for renewable energies, as there are still European wide barriers to entry in frequency-regulation services markets [Bo18]. In Germany it is neither possible to operate on FCR markets with fluctuating energies [Ge18d] nor to use control zones crossing or even cross-national pools of power plants for providing control reserve [Eu17]. With regard to harmonized European power markets and the growing importance of PV and wind energy for the system stability [Bu15], the investigation of the ability of renewable energies performing FCR is a relevant future topic.

2 Related Work

The comprehensive method of this work is a new approach to evaluate ancillary services by aggregated renewable power plants in physical field tests automatically.

The system architecture of a VPP composed of distributed energy resources with the capability to provide grid frequency support has already been shown in [Es17], though the case study results were based on simulations instead of physical field tests. A methodological challenge for developing performance indicators for grid operations is how to translate laws or regulatory frameworks into equations. Requirements analyses based on technical grid codes have been presented in [Dí14], [HSS16], [LER12] and [PVG17],

whereby [LER12] and [PVG17] also develop specific derived key figures to measure the system stability, the availability or the deviations from a certain set point, for example. [LER12] defines KPIs of control reserve provision integrating the dynamics of the response of the balancing provider. However, none of these authors has developed one global KPI that allows to evaluate the performance quality respectively the fulfilment of the grid requirements at a glance. Moreover, the evaluations are either based on simulations [HSS16], [PVG17] or do not consider variable renewable production units [LER12].

Whereas this work proposes a method to evaluate automatically real physical control reserve field tests of renewable power plants within a VPP based on one main KPI which is derived from the technical requirements of the existing grid codes.

The system architecture for the control reserve field tests is presented in section 3.1, and the methodology for the performance evaluation in section 3.2.

3 Proposed Methods

3.1 System Architecture for Building a Transnational VPP

The system architecture of the VPP (see Fig. 1) consists of different major components that are described more in detail below. The central component of the VPP is the control system, which coordinates all FCR processes.

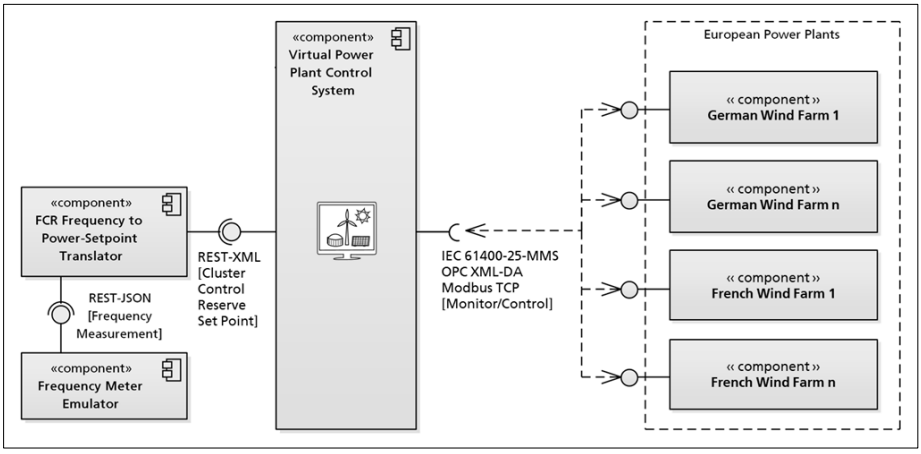


Fig. 1: REstable field test system architecture for FCR

On the customer side, the power plants can be connected via various interfaces. Within the project, the different power plants were connected via IEC 61400-25-MMS, OPC XML DA and Modbus TCP. *IEC 61400-25-MMS* is a communication standard for a

uniform information exchange for monitoring and controlling wind power plants and was developed by the technical committee «Wind energy generation systems» (IEC TC 88) [MTH09]. It is based on IEC 61850, which is used for communication in substation automation. The standard defines the components of a wind turbine in a manufacturer-independent environment and serves to exchange the information provided by the components [Le08]. *OPC XML-DA* was released by the OPC foundation and relies on the OPC DA server based on COM/DCOM technology. Due to some limitations of that technology like platform dependency and communication issues, the OPC foundation decided to further develop the specification based on a SOAP (Simple Object Access Protocol) web service via HTTP (Hypertext Transfer Protocol) [HY10]. *Modbus TCP* has become a standard communication protocol for connecting industrial devices in a vendor-neutral way. It is commonly used in SCADA systems for communication with programmable logic controllers (PLCs) [MNK14]. All three protocols define the data points and the way to communicate. In addition, IEC 61400-25-MMS defines a data model for a uniform information exchange for monitoring and control. This data model is vendor-specific in OPC XML-DA and Modbus TCP.

The architecture used in the project differs from a current FCR architecture. In real operation, each power plant would have its own frequency meter. The active power response would be activated autonomously by each power plant based on local frequency measurement [Sw06], [Co14]. The control system would only be responsible for the distribution of the shares of control reserve provision, which are disaggregated for each participating power plant. In contrast to this, in the architecture for control reserve field tests it is useful to perform the frequency measurement centrally at the control system to avoid installation expenses for several power plants. For the REstable project's field test purpose the system architecture has been modified to the extent that a frequency meter emulator that emulates a frequency signal of the grid is used instead of a central frequency measurement. The values are read in via a csv-file and are provided to the control system via a REST-JSON interface. It has the advantage that a common and defined frequency curve can be used for all power plants and for each test in order to be able to build reproducible and evaluable field tests. For the FCR field tests an FCR bid is set manually in the VPP control system, the amount of which depends on the weather conditions for wind and solar.[3] The conversion of the given grid frequency into an active power target for the tests is performed by a component called «Frequency to Power-Setpoint Translator», which is connected to the frequency emulator via a REST-JSON interface and to the VPP control system via a REST-XML interface. The converter calculates the active power reserve set point based on the frequency using the frequency power curve (see Fig. 2). The calculated amount of target activation is send to the VPP control system. The target active power is disaggregated for the participating power plants by the VPP control system. This disaggregation takes into account the available active power (AAP) and the state of each power plant. The result of the disaggregation is a necessary active power reduction for each power plant, which should be realized to

[3] In the target state of the system architecture, a connected trading tool would be responsible to submit accepted FCR offers to the control system.

provide the required amount of FCR.

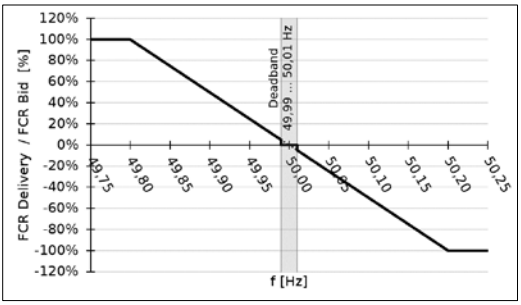


Fig. 2: FCR power frequency characteristic [Ge14]

The following subchapter presents the evaluation methodology for FCR field tests, including the requirements analysis, the development of KPIs and the Python model.

3.2 Performance Evaluation of Control Reserve Supply

Requirements Analysis of FCR

The FCR is activated non-selectively and in solidarity throughout the ENTSO-E network. It has to be provided symmetrically in positive and negative direction and must be available for up to 15 minutes within 30 seconds in case of activation. [Co14], [Sw06]

The following table contains the most relevant requirements for the KPI evaluation that are based on the official German TSO's documents.

ID	Short Title	Description	Source
FCR_01	Provision time slice	The VPP must provide the contracted FCR for the time slice of one week (Mo 0:00 to Su 24:00).	[Bu11]
FCR_02	Provision availability	The VPP must ensure a 100 % availability of contracted power during the whole provision period.	[As03]
FCR_03	Activation time	The VPP must activate the required power reserve within maximal 30 seconds.	[As03]
FCR_04	Duration of activation	The VPP must be able to deliver the contracted FCR for at least 15 minutes.	[As03]
FCR_05	Over- and underfulfillment	The VPP must ensure a maximal overfulfillment of the maximum value of (5 MW, 20 % of the set point) and avoid any underfulfillment.	[Ge13]

Tab. 1: German FCR requirements

The amendment of the FCR requirements as in [Ge18b] leads to some modified and as well some additional requirements, which are presented in Tab. 2.

ID	Short Title	Description	Source
FCR_06	Power change areas	The period after a set point change is separated into a power change area (0...30 s), a transient area (30...90 s) and a stationary area (90...n s).	[Ge18b]
FCR_07	Gradient	The VPP has to activate the FCR evenly, i.e. the first 50 % in the first 15 seconds and the last 50 % linearly in the next 15 seconds.	[Eu17], [Ge18b]
FCR_05	Over- and under-fulfillment	An over- and underfulfillment of 20 % (10 %) of the set point is permitted resp. of 30 % (20 %) is tolerated during the transient (stationary) area.	[Ge18b]
FCR_08	Permitted / tolerated corridor	The VPP must ensure at least 95 % of the values to be within the permitted corridor and maximal 5 % within the tolerated corridor.	[Ge18b]

Tab. 2: Amendment of German FCR requirements

A supporting visualization of the requirements regarding the activation times and the tolerance corridors is illustrated in Fig. 3.

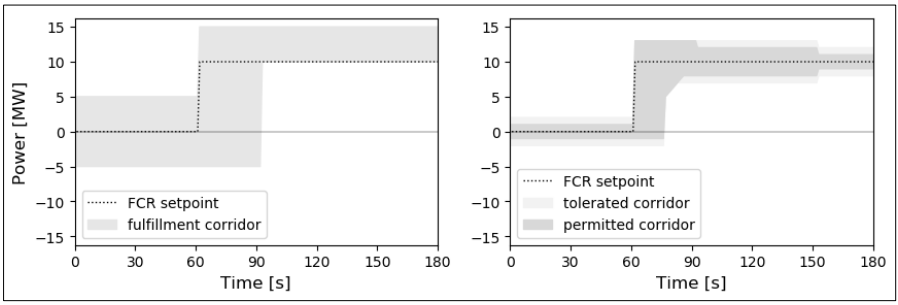


Fig. 3: Limits of fulfilment corridor (left) and limits of permitted and tolerated corridor (right)

Key Performance Indicators

The following time series can directly be exported from the VPP to use them for further calculations.

- Active power $P(t)$
- Available active power $aap(t)$
- FCR bid (provision of ordered control reserve capacity)
  - positive $prov_{ord,pos}(t)$
  - negative $prov_{ord,neg}(t)$
- Set point control reserve (activation call) $s(t)$

- Real time schedule $rts(t)$

$$rts(t) = aap(t) - prov_{ord,pos}(t) \tag{1}$$

- Actual value control reserve (activation) $av(t)$

$$av(t) = P(t) - rts(t) \tag{2}$$

In this work, one exemplary KPI, the activation quality, is presented since it is the best comparable one. The following time series are calculated to further calculate the KPI.

- Acceptance channel

The following formulas for the upper limit $uac_{FCR}(t)$ and lower limit $lac_{FCR}(t)$ of the acceptance channel consider the requirements for control reserve (*FCR_03* in Tab. 1) and are adapted to official public formulas of the German TSOs as in [Ge18a].

$$uac_{FCR}(t) = \max\{s(t-31), \dots, s(t)\} \tag{3}$$

$$lac_{FCR}(t) = \min\{s(t-31), \dots, s(t)\} \tag{4}$$

According to the new prequalification requirements (see *FCR_07* in Tab. 2), the upper limit $uac_{FCR}(t,j)$ and lower limit $lac_{FCR}(t,j)$ of the acceptance channel for FCR are now determined by a reaction time $T_{react} = 15\,s$ and a certain gradient $g(t)$ after each time-stamp t_{change} with a set point change. The equations below refer to the formulas for the calculation of tolerance limits in [Ge18c].

$$g(t) = \frac{|s(t) - s(t-30)|}{30\,s} \tag{5}$$

$$t_{change} \in \{t, s(t) \neq s(t-1)\} \tag{6}$$

For $\forall\, j \in \{t_{change}, \dots, t_{change} + 2 * T_{react}\}$:

$$uac_{FCR}(t,j) = \begin{cases} s(t), & s(t) \geq s(t - 2 * T_{react}) \\ s(t - T_{react}), & [s(t) < s(t - T_{react})] \wedge [j \leq (t_{change} + T_{react})] \\ s(t - 2 * T_{react}) - (j - t_{change}) * g(t), & s(t) < s(t - 2 * T_{react}) \end{cases} \tag{7}$$

$$lac_{FCR}(t,j) = \begin{cases} s(t), & s(t) \leq s(t - 2 * T_{react}) \\ s(t - T_{react}), & [s(t) > s(t - T_{react})] \wedge [j \leq (t_{change} + T_{react})] \\ s(t - 2 * T_{react}) + (j - t_{change}) * g(t), & s(t) > s(t - 2 * T_{react}) \end{cases} \tag{8}$$

- Fulfillment corridor / tolerated and permitted corridor

The upper and lower limit of the acceptance channel is extended by certain tolerances which are defined in requirement number FCR_05 in Tab. 1 and Tab. 2. The calculation of the upper and lower limits of the tolerance corridors is performed by a Python based analysis framework and leads to the visualization in Fig. 3. The formula symbols for the different corridors are summarized in Tab. 3.

Corridor	Upper limit	Lower limit
fulfillment corridor	*ufc(t)*	*lfc(t)*
permitted corridor	*upc(t)*	*lpc(t)*
tolerated corridor	*utc(t)*	*ltc(t)*

Tab. 3: Formula symbols upper and lower corridors

In order to determine the total activation quality, the share of values within the fulfilment corridor respectively within the tolerated and permitted corridor is measured.

- Values within fulfillment corridor / tolerated and permitted corridor

$$av_{fulfillment}(t) = \begin{cases} true, & lfc(t) \leq av(t) \leq ufc(t) \\ false, & else \end{cases} \tag{9}$$

$$av_{permitted}(t) = \begin{cases} true, & lpc(t) \leq av(t) \leq upc(t) \\ false, & else \end{cases} \tag{10}$$

$$av_{tolerated}(t) = \begin{cases} true, & ltc(t) \leq av(t) \leq utc(t) \\ false, & else \end{cases} \tag{11}$$

- Share of values within fulfillment corridor / tolerated and permitted corridor

$$av_{fulfillment,true} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{av_{fulfillment_i}, av_{fulfillment_i} = true\right\} \tag{12}$$

$$av_{permitted,true} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{av_{permitted_i}, av_{permitted_i} = true\right\} \tag{13}$$

$$av_{tolerated,true} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\left\{av_{tolerated_i}, av_{tolerated_i} = true\right\} \tag{14}$$

- Activation quality

The activation quality corresponds to the share of values within the fulfilment corridor. For the new requirements, it is the share of values within the permitted corridor plus

max. 5 % of the measured values within the tolerated corridor (see *FCR_08* in Tab. 2).

$$quality_{old}(\%) = av_{fulfillment,true} * 100 \quad (15)$$

$$quality_{new}(\%) = \left(av_{\text{permitted},true} + \max\{av_{tolerated,true}, 0.05\}\right) * 100 \quad (16)$$

All formulas can be calculated separately for times of positive FCR activation, negative FCR activation and zero FCR activation. This might be interesting to investigate misbehaviour of the power plants in certain situations, because from a technical point of view it is another challenge to provide positive control power compared to negative.

Python Model for Automatic KPI Evaluations

The evaluation in Python can be generated after each field test. The modular structure of the Python model is described below and is illustrated in Fig. 4.

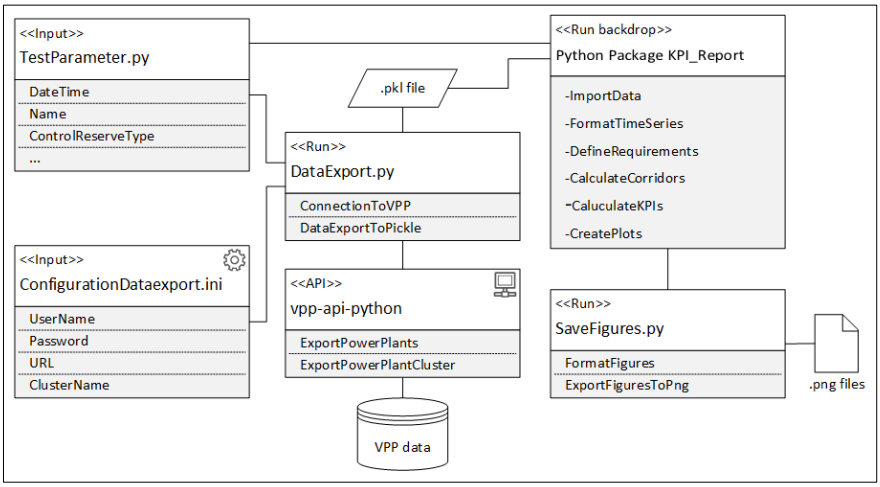


Fig. 4: Python model for the automatic FCR test evaluation

The Application-Programming-Interface (API) *vpp-api-python* allows the access to the VPP data. The configuration file *ConfigurationDataexport.ini* contains connection and configuration details for the VPP data export. In the script *TestParameter.py* some test and wind farm specifics have to be defined by the user. The script *DataExport.py* reads the information of the configuration file and the test parameters. It exports defined time series via the API from the VPP data and safes it in a pickle (.pkl) file which is useful for archiving large amounts of data in a Python conform data format. The different scripts in the Python package *KPI_Report* have functions to import the time series from the pickle file, to convert it to pandas data frames, to calculate the tolerance corridors and the KPIs and to create plots for the visualization of the evaluations and time series. The script

SaveFigures.py activates the Python package *KPI_Report* backdrop and finally exports the produced plots of time series and evaluation diagrams to PNG files.

In the next chapter, a case study for executing and evaluating FCR field tests with application of the described methods is presented.

4 Case Study

4.1 Performing Transnational Control Reserve Tests

For the control reserve tests, the frequency emulator simulates a frequency deviation in the grid. First, a reduction of the grid frequency of 200 mHZ is simulated for 15 minutes (minutes 20 to 35 in Fig. 5) and an increase of the grid frequency of 200 mHz is simulated between minutes 50 and 65. This is analogical to the FCR model protocol that has to be performed in Germany to prove FCR ability [Ge18b]. The stepwise activation of control reserve enables a clear observation of the active power response of the VPP. According to the frequency power characteristic in Fig. 2, a frequency deviation of +/- 200 mHz corresponds to the complete activation of the FCR bid.

An exemplary reaction of the power plants to the FCR bid and the activated FCR is shown in Fig. 5, that visualises the time series active power $P(t)$, available active power $aap(t)$ and the real time schedule $rts(t)$.

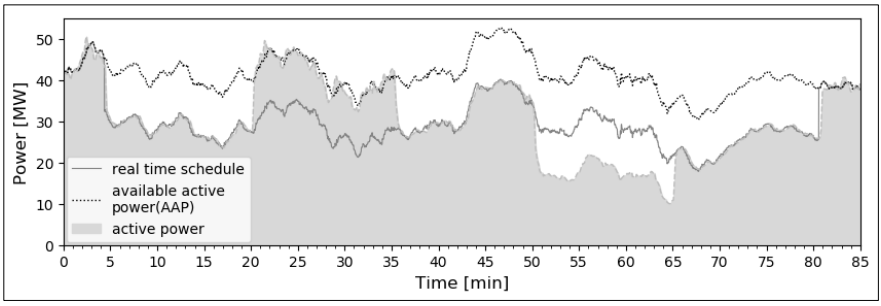


Fig. 5: Physical reaction of a wind farm cluster to FCR set points

The real time schedule always corresponds to the value of the available active power reduced by the amount of the positive FCR bid. The control reserve mode of the power plant cluster is activated about 30 seconds before the start of the FCR bid, so that the active power can follow the real time schedule just in time to fulfil the requirement for 100% availability of provided FCR (see *FCR_02* in Tab. 1). The control system reduces the active power generation by the amount of the positive FCR bid to be able to activate the positive control reserve. During the phase of the simulated frequency drop (rise), the power plant cluster receives the set point signal to power up (down) the active power by

the amount of the positive (negative) FCR bid plus a safety factor. The activated FCR $av(t)$ in positive or negative direction is equivalent to the deviation of the active power from the real time schedule.

The mentioned safety factor is implemented in the VPP as an offset correction function and is necessary for two reasons. First, to avoid any underfulfillment because of fluctuations of the active power. Second to compensate a possibly inaccurate calculation of the AAP that has to be forecasted rollingly by the power plants themselves.

In the following subchapter, one exemplary evaluation of a control reserve field test with the developed KPIs is presented.

4.2 Application of KPIs to REstable Field Tests

In the following evaluations, the offset correction function of the VPP will be disregarded in order to get a clear comparison of the KPIs according to the old and the new requirements without any computational modification of time series. In this example, the power plant cluster receives a set point of 12.5 MW for positive FCR activation and 10.5 MW for negative FCR activation. Fig. 6 shows an extract of the FCR time series with the according different tolerance corridors. The old requirements apply in the left diagram and the new requirements in the right diagram.

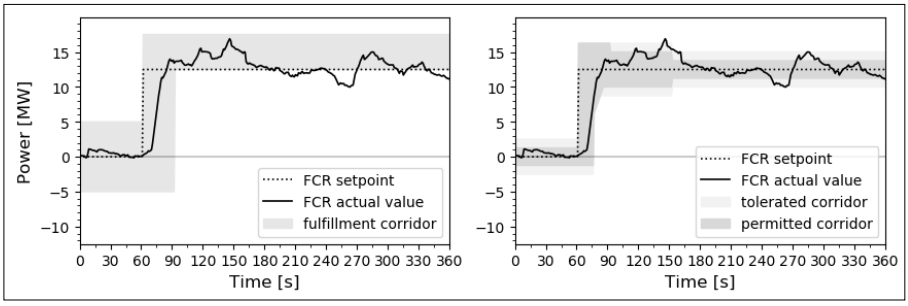


Fig. 6: Limits of fulfilment corridor (left) and limits of permitted and tolerated corridor (right)

One can see clearly the underfulfillment of the required FCR caused by an overestimation of the AAP (left diagram). The permitted and tolerated corridors for both positive and negative deviations of the FCR according to the new requirements (right diagram) lead to the result, that there is no underfulfillment in the illustrated extract.

The activation quality for the whole time slice of the test is presented in Fig. 7, separated into positive and negative FCR and times without FCR call. For the positive and negative FCR, the quality is better according to the new requirements, as they allow both over- and underfulfillment to a certain extent. This leads to an activation quality of 83.94 % according to the old requirements and 90.58 % for the new ones.

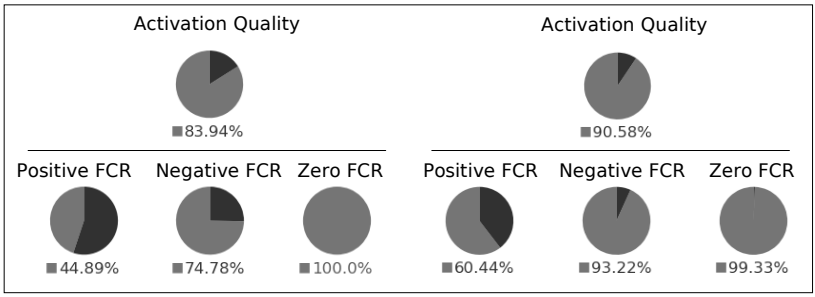


Fig. 7: Activation quality control reserve field test for old (left) and new requirements (right)

5 Discussion and Conclusion

The operated control reserve field tests as part of the REstable project are a success regarding the transnational established ITC infrastructure and the overcoming of regulative and technical obstacles for a centralized control of decentral European fluctuating power plants.

The developed KPIs allow a transparent verification of the fulfilment of control reserve requirements. A quantitative interpretation of the results regarding the activation quality is currently difficult as there are no comparable figures available. Moreover, the project did not claim to draw statistically relevant conclusions from the tests, because the number of tests was limited due to an agreed limited amount of energy losses. Nevertheless, some aspects could be detected which lead to loss of performance in the KPIs. First, a systematic overestimation of the calculated AAP can cause a merely arithmetical underfulfillment of the required FCR value. The offset correction of the VPP can compensate this theoretically, but it is not an officially proven procedure yet. Second, some general technical aspects can lead to slower reaction times of the power plants. For example, the interval for the acceptation of set point signals by the power plants has to be adjusted as short as possible, as current default values are up to 60 seconds to protect the power plants. Moreover, the communication chain in this system architecture from the central frequency emulator via the power target converter via the aggregator to the power plants can lead to latencies of a few seconds.

The further development of the technical requirements for control reserve seems to be more fitting to the control behaviour and the variability of renewable energies. According to the new requirements, a deviation from the set point is tolerated in both positive and negative direction and a distinction is made between a transient and a stationary area after each set point change. Moreover, the results of the field test show a better activation quality for the new requirements.

As a conclusion, it is possible to perform FCR of good quality with fluctuating energies aggregated transnationally within a VPP. A hundred percent fulfilment of the TSO's

requirements on FCR is still not possible. Moreover, some regulations have to be amended to open the FCR market for fluctuating energies, such as the lengths of the bidding time slices and the lead times or the permission for control area crossing aggregation of power plants, for example.

Acknowledgments

The work reported in this paper was part of the European research project 'REstable - Improvement of renewables-based system services through better interaction of European control zones' within the ERA-Net Smart Grid Plus program. The authors gratefully acknowledge the contribution of the BMWi (DE), the ADEME «Investissements d'Avenir» (FR) and the Foundation for Science and Technology (PT).

Bibliography

[Ac15] Ackermann, T. et al.: Integrating Variable Renewables in Europe. Current Status and Recent Extreme Events. In IEEE Power and Energy Magazine, 2015, 13; pp. 67–77.

[ADA18] Attya, A. B.; Dominguez-Garcia, J. L.; Anaya-Lara, O.: A review on frequency support provision by wind power plants. Current and future challenges. In Renewable and Sustainable Energy Reviews, 2018, 81; pp. 2071–2087.

[As03] Association of German Grid Operators (VDN): TransmissionCode 2003 - Anhang D1, 2003.

[Bo18] Borne, O. et al.: Barriers to entry in frequency-regulation services markets. Review of the status quo and options for improvements. In Renewable and Sustainable Energy Reviews, 2018, 81; pp. 605–614.

[Bu11] Beschluss BK6-10-097. Festlegung zu Verfahren zur Ausschreibung von Regelenergie in Gestalt der Primärregelung, 2011.

[Bu15] Bundesministerium für Wirtschaft und Energie (BMWi): Ein Strommarkt für die Energiewende, 2015.

[CMK18] Camal, S.; Michiorri, A.; Kariniotakis, G.: Optimal Offer of Automatic Frequency Restoration Reserve From a Combined PV/Wind Virtual Power Plant. In IEEE Transactions on Power Systems, 2018, 33; pp. 6155–6170.

[Co14] Consentec: Beschreibung von Regelleistungskonzepten und Regelleistungsmarkt. Studie im Auftrag der deutschen Übertragungsnetzbetreiber, 2014.

[Dí14] Díaz-González, F. et al.: Participation of wind power plants in system frequency control. Review of grid code requirements and control methods. In Renewable and Sustainable Energy Reviews, 2014, 34; pp. 551–564.

[Es17] Essakiappan, S. et al.: Dispatchable Virtual Power Plants with forecasting and decentralized control, for high levels of distributed energy resources grid penetration: 2017

IEEE 8th International Symposium on Power Electronics for Distributed Generation Systems (PEDG). IEEE, 2017; pp. 1–8.

[Eu17] VERORDNUNG (EU) 2017/1485 DER KOMMISSION zur Festlegung einer Leitlinie für den Übertragungsnetzbetrieb. System Operation Guideline, 2017.

[Fr14] Fraunhofer IWES: Regelenergie durch Windkraftanlagen, 2014.

[Fr17] Fraunhofer IWES: Regelenergie durch Wind- und Photovoltaikparks, 2017.

[Ge13] German TSOs: Rahmenvertrag über die Vergabe von Aufträgen zur Erbringung der Regelenergieart Primärregelleistung, 2013.

[Ge14] German TSOs: Eckpunkte und Freiheitsgrade bei Erbringung von Primärregelleistung, 2014.

[Ge18a] German TSOs: Info SRL-Abrechnung. Marktinformation Anpassung der Abrechnungsbedingungen für Sekundärregelarbeit, 2018.

[Ge18b] German TSOs: Präqualifikationsverfahren für Regelreserveanbieter (FCR, aFRR, mFRR) in Deutschland. PQ-Bedingungen, 2018.

[Ge18c] German TSOs: Leitfaden zur Bestimmung von Regelleistungsistwerten, 2018.

[Ge18d] German TSOs: Leitfaden zur Präqualifikation von Windenergieanlagen zur Erbringung von Minutenreserveleistung im Rahmen einer Pilotphase, 2018.

[HSS16] Hess, T.; Schegner, P.; Schmidt, M.: Studies on provision of ancillary services by distributed generation units and storage devices. In (Gubina, A. F. Ed.): IEEE PES Innovative Smart Grid Technologies, Europe. October, 9-12, 2016, Ljubljana. IEEE, Piscataway, NJ, 2016; pp. 1–6.

[HY10] Huiming, L.; Yao, Y.: The research and development of OPC XML-DA Server based on web service technology: 2010 2nd International Conference on Advanced Computer Control. IEEE, 2010; pp. 472–475.

[Le08] Lee, J.-H. et al.: IEC 61400-25 interface using MMS and web service for remote supervisory control at wind power plants: 2008 International Conference on Control, Automation and Systems. IEEE, 2008; pp. 2719–2723.

[LER12] Lobato, E.; Egido, I.; Rouco, L.: Monitoring frequency control in the Turkish power system. In Electric Power Systems Research, 2012, 84; pp. 144–151.

[MNK14] Mohammadzadeh Fakhr Davoud, A.; Navi, M. G.; Kanani, S. K.: Online monitoring of gas turbine power plant using modbus/TCP: 2014 Smart Grid Conference (SGC). IEEE, 2014; pp. 1–5.

[MTH09] Min-Jae Seo; Tae-o Kim; Hong-Hee Lee: Implementation of web services based on IEC 61400-25 for wind power plants: ICROS-SICE International Joint Conference 2009.

[PVG17] Pinceti, P.; Vanti, M.; Giannettoni, M.: Technical KPIs for microgrids: 2017 IEEE International Systems Engineering Symposium (ISSE). IEEE, 2017; pp. 1–7.

[Sw06] Swider, D. J.: Handel an Regelenergie- und Spotmärkten. Dissertation, 2006.

Using grid supporting flexibility in electricity distribution networks

Immanuel König [1], Erik Heilmann [2], Janosch Henze [3], Klaus David [1], Heike Wetzel [2], Bernhard Sick [3]

Abstract: The electrical grid is facing several challenges. On the energy generation side is the decentralized power generation in solar parks, wind parks, or residential solar panels, which all result in a time variable power generation. They may generate power while it is not needed. On the load side, there are the challenges of controllable loads and the electromobility with its high demands on the grid. These loads may need power when it is not available or cannot be transmitted by the grid. In addition to these technical challenges, there is the political will in Germany both towards renewable energy generation and towards a smart grid. The orchestration of controllable loads with variable power generation could alleviate these problems. Loads could be orchestrated to use power when it is generated. In this paper two approaches of how to orchestrate the controllable loads considering the variable power generation are described: a centralized and market-driven approach from the research project C/Sells and a decentralized approach from the research project LAGE-EE.

Keywords: renewable energy, smart grid, flexibility, energy trading

1 Introduction

The electrical grid should perform reliable and efficient, especially for the optimal use of the already installed grid. Additionally, there are new challenges for the grid. Due to the uprising use of renewable energy sources, the location of the sources can be scattered all over the area and thus all over the grid [FH19]. Especially in Germany, many solar arrays are built on domestic house roofs. And also, solar parks and wind parks are not necessarily built in strategic positions for the grid, but rather where the wind or the necessary space is available. Due to the nature of these generators, the amount of produced energy can vary quickly, mainly related to the weather. Thus, the energy may be available when it is not needed or where it is not needed, or the generation of energy may stop weather related while it is still needed.

On the energy consumption side are also challenges. Some consumers are, to some degree, flexible in the amount of energy needed and the time when the energy is needed. To coordinate the energy generation and consumption an act [DB16] has been passed in Germany, giving the possibility to control some energy consumers and this in a „smart“

[1] University of Kassel, Chair of Communication Technology, david@uni-kassel.de
[2] University of Kassel, Institute of Economics, address, heike.wetzel@uni-kassel.de
[3] University of Kassel, Chair of Intelligent Embedded Systems, address, bsick@uni-kassel.de

doi:10.18420/inf2019_70

way. Flexible energy loads raise the challenge of the best regulation mechanism to control them. In addition, an increase of electromobility can be seen [NP18]. This will also significantly increase the load in the grid, especially the maximum power required at times of loading many vehicles. An upgrade of the existing grid, allowing to transport larger electric powers, could solve the described challenges. Nevertheless, each upgrade of the grid is expensive. Therefore, the optimal use of the existing grid should be considered.

A solution to the described challenges can be the smart control or orchestration of both sources and consumers of energy. But then, the question for the best regulation mechanism remains. In this paper two different approaches are presented. Both are developed in research projects at the University of Kassel. In the first research project C/Sells [CS19], a trading-based approach is developed and investigated, to cover not only the technical aspects but also the monetary aspects. This trading aims to optimize the overall energy situation given by time-varying energy production and potentially time-varying energy consumption. So, i.e., energy consumers could be motivated to shift their energy consumption due to financial reasons. This project focusses on the inclusion of loads on a corporate level like cooling warehouses and process ovens. The second research project LAGE-EE [LA19] concentrates on flexible energy consumers on a domestic scale. In this project, a field test is done where heat pumps are controlled by the local electrical parameters of the grid. While the first approach focuses on a centralized and market-driven orchestration, the second approach investigates a decentralized and locally controlled approach.

The remainder of the paper is as follows: First, an overview of the problems of an electrical grid is given. Afterwards, the two approaches from the two research projects are explained in detail. Finally, a conclusion is given.

2 Detailed description of the issues in electrical grids

In this section, an overview of the German grid structure is given. Also, the issue of congestions is explained. This is the basis for the solutions investigated in the R&D projects C/Sells and LAGE-EE. These solutions will then be explained in Section 3 and 4, respectively.

2.1 Structure of the grid

The grid in Germany can be divided into two layers: the transmission network (the upper layer in Figure 1) and the distribution network (the lower layer in Figure 1). The transmission network is used to deliver the energy over long distances from the energy source to the distribution networks. Transmission networks can also be used to transfer energy from a distribution grid to another distribution grid. And they can even cross international borders, to exchange energy between countries.

Distribution networks are the second stage of the grid. They are operated in medium, high and low voltage and connect the transmission networks with the local grids and the residential houses. Almost all wind farms (96%) [BW19], and many solar farms as well as companies with higher energy needs are connected to the medium-high voltage grid. At the low voltage grid, the residential houses and small companies are connected via a local transformer. Also, at this voltage level, the typically large number of residential solar arrays are connected. On the other hand, the large powerplants are only connected to the first layer, the transmission grid.

The traditional energy flow in the German grid was top-down, starting from the power plant over extended transmission networks and ending in the residential houses or the industry. In this traditional flow all the generators would only be in layer 1 in Figure 1 and all the loads in layer 2.

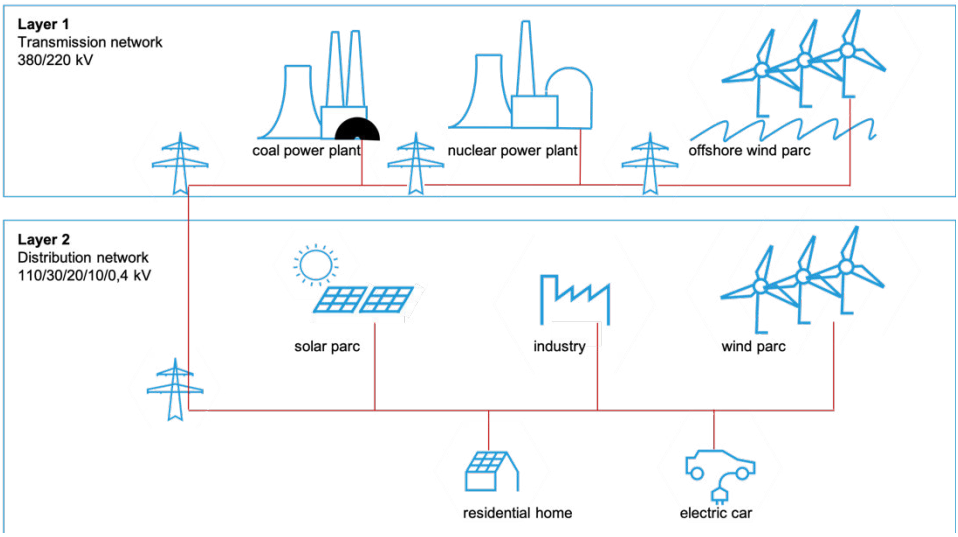


Figure 1: German grid, energy flow and layers [VU15]

2.2 Congestions

The traditional setup of the grid in Germany was load oriented. The energy was produced in large powerplants and distributed from the transmission networks over the distribution networks to the local consumers. The distribution of load over time was approximately known, and the energy generation was planned, scheduled, and controlled to the needs of the customers.

The traditional grid setup has changed over the last years. Many private households have installed solar arrays on their roofs. In addition to this, wind and solar parks have been

installed in many locations. They all feed energy into the grid. Unfortunately, the amount of energy feed into the grid is depending on the weather and the time of the day and year. Also, a change in energy needs is expected due to the rising number of electro mobiles and heat pumps. These two developments lead to the risk of energy congestions. Energy congestion means, that the grid is not able to support as much energy transport. Congestions can occur either on the energy consumption side, as well as on the energy production side.

Congestions induced by the energy production side occur when there is too much energy produced but not needed in the local grid, and it cannot be fed back into the upper network layers. This is illustrated in Figure 2 at 'House 2'. The voltage rises over the upper border V_{max}. To stabilize the grid the generation of energy is reduced by the solar inverter of House 2 until the grid is in the allowed electrical parameters again.

Congestions induced by the energy consumption side occur when not enough energy can be transported to the consumer. A scenario covering this hypothetical situation is the homecoming of employees in the evening with their electric cars. They all connect their vehicles to the grid and want to charge them. If many residents of a single street or a village arrive at a similar time, congestion of energy may occur, because the installed grid may not be able to deliver the energy needed to charge all vehicles fast and at the same time. A Congestions of this kind is illustrated in Figure 2 under the SME. As a result of such a situation the voltage in this location of the grid drops under the minimum V_{min}.

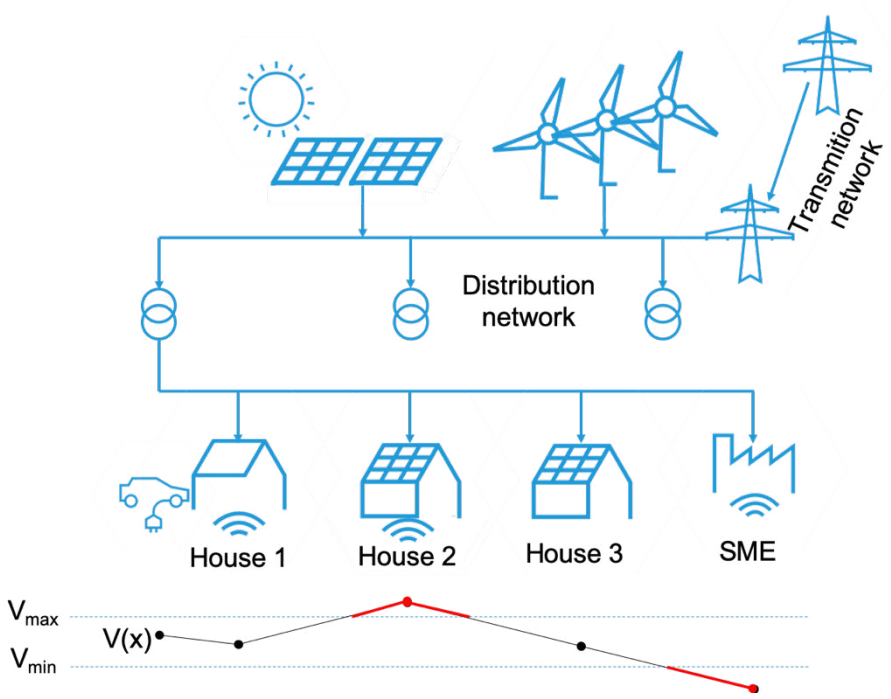


Figure 2: Congestions in the distribution Network

2.3 Grid expansion

One way to solve this issue of congestions is an upgrade of the power grid. The power transmission capacities of the grid mainly depend on the capacities of the installed technical infrastructure like power lines and transformers. A change of the installed infrastructure to an infrastructure with a higher capacity can solve the issue of congestion caused by too much power generation or too much power consumption. However, an upgrade of the already installed technical infrastructure has to be considered carefully because of the high costs resulting from it. Besides this, it would also be uneconomical to install power lines with a larger capacity when the full capacity is only used rarely. Most of the time it would only be used as it is already used today, and there is no need for a change during this time.

2.4 The use of flexibilities to stabilize the grid

The German government agency of regulation (Bundesnetzagentur) defines flexibilities in the context used in this paper: "On an individual level flexibility is the modification of generation injection and/or consumption patterns in reaction to an external signal (price signal or activation) in order to provide a service within the energy system. The parameters used to characterize flexibility include the amount of power modulation, the duration, the rate of change, the response time, the location etc." [BN17]

Flexibilities can be used to stabilize the grid. They can be controlled to generate more or consume less power when there is a congestion of energy generation. And when there is an oversupply they can be controlled to generate less or consume more power. This can be done in many locations in the grid simultaneously, depending on the local requirements of the grid.

The control of flexibilities can be done by different approaches. Two approaches are described in the next two sections.

3 The approach of regional flexibility markets: C/Sells

The research project C/Sells investigates on the future energy grid setup of Germany. One particular focus is the use of flexibilities to solve congestions in the grid. The flexibilities which may help to solve a congestion have to be chosen out of the available flexibilities. In this research project the flexibilities get chosen via a market-based mechanism. The underlying market mechanisms are investigated, and several mechanisms are tested. This investigation is supported by a test done in a simulation and a field test.

The flexibilities considered for a market in this project are on a corporate level like cooling warehouses and process ovens. However, the same underlying principals can be used to enable other participant on a smaller or larger scale.

In the next subsection the market platform is described. Then the structure of the market and the traded products are described in more detail. In the last subsection the forecast mechanism to predict congestions is described. This forecast is needed to trigger the allocation of flexibilities and therefore the market process.

3.1 Market based use of flexibility

The general idea of a flexibility market is the procurement of a required flexibility via market mechanisms. This is done to enable a transparent and non-discriminatory allocation of a flexibility. A platform is needed for the market. This market platform in connection with its participants is shown in Figure 3 .

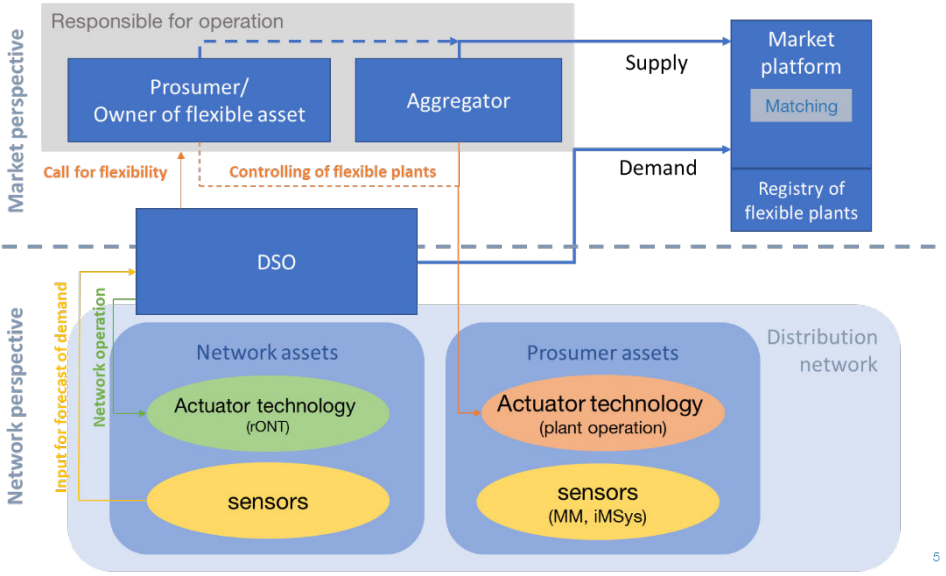


Figure 3: Concept of the market platform

The lower part of Figure 3 shows the network perspective. The classic network operation comprises the handling of active network infrastructure and will be based more on different kinds of sensors (e.g., smart meters) in the future. The data of these sensors can be used to monitor and predict the state of the network (see also Section 3.3). With this information, the network operator can identify problems and can determine the demand for flexibility. The demand for flexibility gets transmitted to the market platform.

The market platform is also the place, where owners of flexible assets can bid their flexibility. Flexible assets can be, for example, households with battery storage, industrial companies with flexible demand, flexible power plants, or aggregators that bring a pool of different power plants to this market.

The demand for flexibility and the bidden flexibilities are matched on the market platform. The output of the market platform is a list of flexibilities that can solve the expected problem with the lowest possible cost.

The network operator can decide if and when he actually wants to use a flexibility from the list of flexibilities. If the predicted situation occurs, he can request a flexibility from the list. The activation of a particular infrastructure on the other hand, is the responsibility of the supplier of the flexibility.

3.2 Structure of the market and traded products

We also implemented a prototypical regional flexibility market for a sub-network of the project partner EnergieNetz Mitte, a DSO (Distribution System Operator) in northern Hesse. The market processes are structured, as shown in Figure 4.

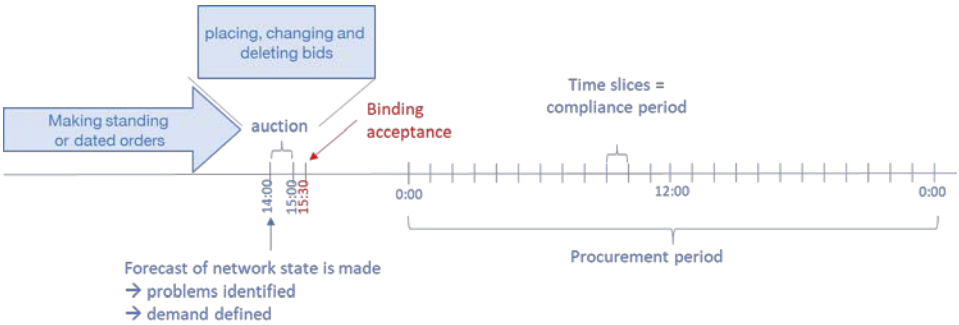


Figure 4: Structure of the market process

The trading process is organized in a day-ahead auction, which takes place between 2 and 3 pm. Until the start of this auction, the network operator must define the demand for flexibility. Suppliers can place, change, and delete their bids until the end of the auction. It is also possible to automate the placement of bids by placing automatically repeating standard bids. After the auction, the winner is determined, and a list of the relevant bids is generated. This list is the output of the market.

The traded products of our market contain the option for the system operator to request the flexibility if actually needed. Therefore, the decision of activation is delayed to the trading of flexibility. One bid of a flexibility product belongs to one hour of the next day, because the trading is organized in a day-ahead auction. If traded, the flexibility supplier must maintain the flexibility in the auctioned time slot. The activation of the flexibility is realized in a short time, e.g., 2 hours, before delivery.

We designed various products that aim to fulfill different requirements for the demand side as well as the supply side. The products of our market are defined in this paragraph and summarized in Figure 5. The main product is a positive or negative power. This is the standard product when talking about flexibility. It leads to the adoption of the production or consumption schedule of the supplier. In connection with this product, we designed three different 'quota' products for different technologies (PV, wind, e-mobility).

'Quota' products do not contain the change of a specific value of electric power, but the limitation of production or consumption of these assets, in contrast to the standard power products. These products respect the difficulty of predicting the production or consumption of individual small plants. At last, we take into account the possibility of solving voltage problems with reactive power.

To solve a congestion problem the location of the flexibility is crucial. The more fare away it is from the location of the problem, the less useful it will be to solve it. Therefore every product specification contains a location information. The location Information is not position on the surface of the earth like in GPS, but rather the information to which knot in the net the flexibility is connected. This distance is also used for the valuation of a specific offer.

<table>
<tr><td rowspan="2">Underlying problem</td><td colspan="3">(power) congestion</td><td colspan="3">Voltage problem</td><td rowspan="2">Superior</td></tr>
<tr><td colspan="6">Can be solved with</td></tr>
<tr><td>Technical good</td><td colspan="4">Power</td><td colspan="2">Reactive power</td><td rowspan="5">Standardized</td></tr>
<tr><td>Direction</td><td colspan="2">+
(Generation)</td><td colspan="2">-
(Consumtion)</td><td>+
(capaci
tive)</td><td>-
(induc
tive</td></tr>
<tr><td>Activation condition</td><td colspan="6">Secure maintenance with short term activation</td></tr>
<tr><td>Predictability</td><td>Exact defined</td><td colspan="2">Quota
PV Wind</td><td>Exact defined</td><td>Quota
E-Mob</td><td colspan="2">Exact defined</td></tr>
<tr><td>Time slices</td><td colspan="6">24 One-hour-time slices</td></tr>
<tr><td>Individual local component</td><td colspan="6">Defined with grid node</td><td>Technical based</td></tr>
<tr><td>Prices</td><td colspan="6">Asking price
(€/kW, €/kWh)</td><td>Indi-vidual</td></tr>
</table>

Figure 5: Traded products on the C/Sells market

3.3 Forecasts

In regional energy markets, knowing future power grid states is crucial. Typically, the market operator uses information about future grid states to trigger an auction period. After the auction, a plan on how to balance out possible faulty grid states is created. Before the actual need to implement the balancing plan, another forecast for the power grid state is triggered. This forecast, also called short term forecast, is performed with more current information about the power grid, therefore, also more accurate about the grid state prediction.

Hence, for regional energy markets, we need two types of forecast:

- For planning: long term forecast, 12h to 36h into the future.

- For implementation: short term forecast: up to 6h into the future.

The power grid state is dependent on each individual state of the components in the power grid.

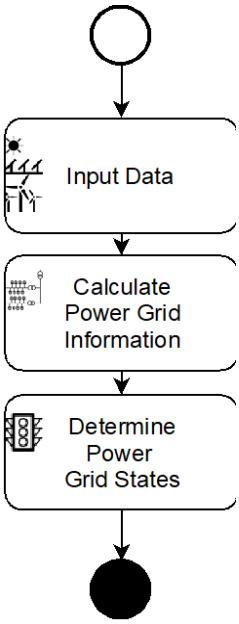


Figure 6: The data analysis chain to obtain power grid states

Mainly the power grid state is obtained after calculating the load flow calculation for the specific grid. Afterward, the results of the load flow allow determining the capacity for each component. Based on the capacity the BDEW (Bundesverband der Energie- und Wasserwirtschaft) [BD01] in Germany proposed limits to define different power grid states.

To be able to forecast the power grid state, we can forecast each of the outputs of three blocks seen in Figure 6:

1. energy production and consumption,
2. load flow results, or
3. grid state estimation.

Each of these approaches increases the difficulty of the problem, as the machine learning algorithm needs to abstract more information about the power grid.

E.g., for Approach 2 the load flow calculation, as well as the power grid layout, needs to be abstracted, and Approach 3 needs to abstract the maximum capacity of the components within the power grid.

Approach 1:
Forecasting energy production and consumption is a straight forward solution to obtain information about the future power grid states. The challenge in this approach is that the number of different forecast models, as an individual forecast for each production and consumption node in the power grid is needed.

Approach 2:
Forecasting the load flow results is a challenge similar to forecasting the energy production and consumption. Each component in the power grid, e.g., each line or transformer, needs to be forecasted, increasing the number of individual forecasts. Additionally, the load flow calculation is added as a preprocessing step.

Approach 3:
Forecasting the state of the power grid is as challenging as forecasting the load flow results. If we use a traffic light system as proposed by BDEW [BD01], the actual forecasting problem changes to a classification problem with three labels.

4 A decentralized control of flexibilities – LAGE-EE

LAGE-EE is a funded research project with a focus on the local management of loads to keep the grid stable. The local management can be useful when there are many residential solar arrays installed in an urban area. The solar arrays can increase the voltage level to a point where the attached solar inverter limits the energy generation because no more energy can be transferred into the grid without violating the stability. On the other hand, there are controllable loads such as heat pumps or electric water heaters. The approach in LAGE-EE is to control such loads locally based on the grid parameters.

The parameter used in LAGE-EE is the local voltage of the grid. A controller has been developed, which is installed in the residential home. It measures the voltage, and when it is during a certain time interval in one of the narrow bands close to the min and max borders of the voltage, the regulation algorithm starts to control the loads.

The loads controlled in the project are mainly heat pumps in residential homes. Heat pumps convert electrical and thermal energy into heat. They achieve a much better efficiency (like a lever) by additionally using temperature differences between inside and outside a building or tapping into the earth. In principle they can also be used for cooling. Here heating is further considered. This heat is used for heating and hot water. Usually, a hot water storage tank is attached to the heat pump, sometimes also one for the heating. And the building itself can also store thermal energy.

Thermal storages can be used to be more flexible in energy consumption. The thermal storages are usually operated within an upper and a lower border. If the room temperature or the temperature in the storage tank approaches the lower border the heat pump is started. When the temperature reaches the upper border the heat pump is stopped. However, the borders can usually be shifted slightly up or downwards. And the heat pump can also be started when the temperature is well between the borders, although this might be unusual. This border and timing shifts enable the heat pump to be more flexible in energy consumption.

Modern heat pumps have already a self-optimized energy consumption function. When solar arrays on the roof produce energy and the energy cannot be feed into the grid because of regulation issues, the heating is started, or thermal storages are directly heated with the excess energy. However, this is only done in one household, not considering the neighborhood. The approach of LAGE-EE comprises a whole neighborhood of several houses and heat pumps, this for the case that locally generated energy gets locally consumed.

The approach in this subsection also helps with the aspired growth rate in number as well as usage of heat pumps. This is part of the political agenda for the use of renewable energy in Germany. Heat pumps are one way to use renewable energy. Thus, many heat pumps get installed into the grid, where traditionally the heating in Germany is done with oil and natural gas [BU15]. The grid is able to supply the heat pumps, as their electric requirements are in within the specifications of the domestic electrical supply. Nevertheless, there can be a problem when several heat pumps start at a similar time in one part of the grid. The controller investigated in LAGE-EE can also help to prevent this problem as the starting of the devices is shifted until the grid is within its specific borders. When the grid is in this state, there is enough electrical energy to start several heat pumps.

The monetary aspects of this approach are also investigated in LAGE-EE. First, the electrical grid in a certain area was simulated. Scenarios were developed for different kinds of future expansions in solar arrays and heat pumps. These scenarios were used in the simulator to calculate the amount of time the grid is not in its specific borders, and the amount of time the energy generation from the solar arrays had to be limited. Afterwards, the potential of buildings and water storage tanks to store energy were investigated. The additional energy losses due to the higher upper border (which also produces higher losses in storages) were simulated. In this simulation also different kinds of buildings were simulated. Also, the avoided grid expansion was considered. Thus, the pricing for the electrical flexibility could be calculated. Models of a dynamic pricing and a flat fee were investigated.

5 Conclusion

The energy generation in Germany is changing towards renewable energies. This changes also the usage of the electrical grid and raises some challenges. Power is no longer only generated in some big plants but rather scattered all over the country. The part of the grid built to supply residential homes is now used to feed energy from PV power plants back into the grid. Wind and solar parks are built which have a fluctuating power generation, depending mainly on the weather. The residential heat and hot-water supply changes to electricity. And the use of electric cars increases, which have a demand for high load currents over a short period.

The use of flexible generation and loads can alleviate several of these challenges. In this paper, two approaches to control flexibilities have been presented, both based on research projects of the University of Kassel. The first approach was using a market mechanism that leads to a transparent and non-discriminatory allocation of flexibilities to be used to stabilize the grid. This new product changes the traditional approach, where only power generation was traded on stock markets. Now also flexibilities can be traded. This may be especially useful for industries utilizing processes that can be shifted in time and also for aggregated smaller loads. In this project also a prototypical regional flexibility market is developed which is also put to a field test.

The second presented approach aimed at a more local stabilization of the grid. It is working decentralized by only measuring electrical properties of the grid at residential homes. The approach is controlling loads like heat pumps based on the local electrical values of the grid. A field test is done in this project to evaluate the simulation-based findings and the regulation algorithm of this approach. Although monetary aspects like the use of excess energy, prevented grid updates and cost increases due to energy losses are considered, the algorithm itself is not considering the price of the energy.

Both approaches can solve some upcoming challenges of the electrical grid by using new financial incentive systems and smart control technologies of the grid. However, future investigations will have to ensure the reliability of these approaches in order to adapt the legal framework to enable such approaches.

6 Literature

[FH19] Fraunhofer ISE: Photovoltaics Report, https://www.ise.fraunhofer.de/content/dam/ise/de/documents/publications/studies/Phot Photovolt-Report.pdf. March 2019.

[DB16] Deutscher Bundestag, Messstellenbetriebsgesetz vom 29. August 2016 (BGBl. I S. 2034), das durch Artikel 15 des Gesetzes vom 22. Dezember 2016 (BGBl. I S. 3106) geändert worden ist, §33. http://www.gesetze-im-internet.de/messbg/index.html, Berlin, 2016.

[NP18] Nationale Plattform Elektromobilität (NPE): Fortschrittsbericht 2018 – Markthochlaufphase. Gemeinsame Geschäftsstelle Elektromobilität der Bundesregierung (GGEMO), Berlin, May 2018.

[LA19] Project Webpage: LAGE-EE - Lastverschiebungspotentiale von Gebäuden für Strom aus erneuerbaren Energien, www.uni-kassel.de/eecs/iteg/forschung/aktuelle-projekte/LAGE-EE.html, May 2019.

[CS19] Project Webpage: C/Sells - Großflächiges Schaufenster im Solarbogen Süddeutschland, https://www.csells.net/de/, May 2019.

[BW17] BWE Bundesverband WindEnergie: Netze, www.wind-energie.de/themen/netze/, 2019.

[VU15] VKU Verband kommunaler Unternehmen e.V.: Das deutsche Stromnetz, www.vku.de/presse/grafiken-und-statistiken/energiewirtschaft/das-deutsche-stromnetz/, May 2015.

[BN17] Project Webpage: C/Sells - Großflächiges Schaufenster im Solarbogen Süddeutschland, www.csells.net/de/, May 2019.

[BU15] Bundesnetzagentur für Elektrizität, Gas, Telekommunikation, Post und Eisenbahnen: Flexibility in the electricity system. Bonn, 2015.

[BD01] BDEW Bundesverband der Energie- und Wasserwirtschaft e.V.: Smart Grid Traffic Light Concept. Format-Verlag, Berlin, P. 14, 2015.

Training of Artificial Neural Networks Based on Feed-in Time Series of Photovoltaics and Wind Power for Active and Reactive Power Monitoring in Medium-Voltage Grids

Marcel Dipp[1] Jan-Hendrik Menke[2] Sebastian Wende - von Berg[3] Martin Braun [4]

Abstract: Today, there is already a significant injection of renewable energies at the medium-voltage level, which requires the use of reliable monitoring methods. In addition to tracking electrical parameters such as line current or bus voltage magnitudes, precise knowledge of the active and reactive power feed-in is becoming increasingly relevant in order to provide the necessary information for optimization strategies at higher voltage levels. For this reason, we have developed a method to monitor the active and reactive power for the medium-voltage level with very low measurement density, which is based on artificial neural networks (ANN). The actual training of ANN is accomplished with photovoltaics (PV) and wind feed-in time series based on real weather data to ensure realistic monitoring of the injection. The presented method is applied to a German medium-voltage grid to evaluate the estimation accuracy.

Keywords: artificial neural networks; active and reactive power monitoring; time series

1 Introduction

The German electricity grid is already subject to an immense feed-in from renewable energy sources in 2019. More than 90 % of the 106.61 GW [Bu19] installed wind and photovoltaic generation capacity is connected to the distribution grid. The feed-in situation will become even more dramatic in the future, as the sum of installed systems will continue to grow considerably over the next decades according to forecast scenarios [DEA13]. This rapid expansion requires knowledge of the current grid condition, especially for the highly fluctuating and time-dependent injections of photovoltaics (PV) and wind energy.
In the past, grid congestion induced by current or voltage did rarely pose problems, so there was no requirement to coordinate distributed energy resources (DER) in medium or low-voltage systems. With the expansion of volatile renewable energy sources and a changed load behavior, the monitoring of the medium-voltage level has become a prerequisite for

[1] University of Kassel, Department of Energy Management and Power System Operation, Wilhelmshöher Allee 71 - 73, D-34121 Kassel, Germany marcel.dipp@uni-kassel.de
[2] University of Kassel, Department of Energy Management and Power System Operation, Wilhelmshöher Allee 71 - 73, D-34121 Kassel, Germany jan-hendrik.menke@uni-kassel.de
[3] University of Kassel, Department of Energy Management and Power System Operation, Wilhelmshöher Allee 71 - 73 and Fraunhofer IEE, Königstor 59, D-34121 Kassel, Germany sebastian.wende-von.berg@uni-kassel.de
[4] University of Kassel, Department of Energy Management and Power System Operation, Wilhelmshöher Allee 71 - 73 and Fraunhofer IEE, Königstor 59, D-34121 Kassel, Germany martin.braun@uni-kassel.de

doi:10.18420/inf2019_71

the grids of the future. Whereas in the transmission grid level the State Estimation (SE) [AGE04] has been state of the art for decades, the minimum number of measuring devices $m_{\min}$ at the medium-voltage level is not sufficient to calculate the complete state of the grid using the traditional SE based on least weighted squares (WLS SE). When using the WLS SE method, missing measurements in medium-voltage grids have to be reconstructed by pseudo measurements. Furthermore, the installation of additional measurement equipment including the information and communication technology causes significant costs for grid operators.

We present a method based on artificial neural networks (ANN) that can provide a sufficiently accurate estimate of active and reactive power injection as they have been trained on feed-in time series. By applying this type of training, the ANN learns various feed-in characteristics based on historical weather data and thus increases the estimation accuracy, especially when estimating feed-ins of large PV and wind power plants. The parameters voltage magnitude and line current are also considered in the studies. Section 2 briefly outlines our comprehensive preceding studies of the presented approach, on which a comparison between the state estimation with ANN and the WLS SE with pseudo measurements has already been evaluated [MBB19]. In Section 3 the paper gives a short insight into the ANN architecture. Section 4 shows how the training of the ANN is performed via feed-in time series and the scenario generator already introduced in [MBB19]. Section 5 highlights the electrical grid used for the validation. Section 6 focuses on the results of the simulation and the achieved monitoring accuracy for active power P, reactive power Q, voltage magnitude V, and line current I.

2 Related Work

The methods covered in this paper are based on detailed preliminary studies and simulations, which are described in [MBB19]. There, an approach for ANN-based grid monitoring was presented. The basic features of the general methodology are shown in Fig. 1. It was demonstrated that the method exceeds the limitations of other existing ANN methods for grid

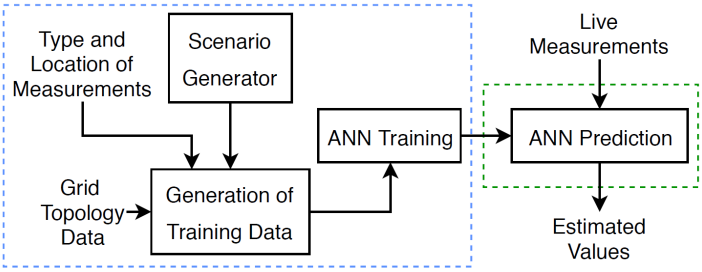


Fig. 1: Flow chart of the scheme: The left side represents the preparation phase, while the right box shows the operating phase, in which live measurements for the trained ANN are used to predict the target values. [MBB19]

monitoring [IG12] [OWM13] [OM14], which are not applicable for electrical distribution grids with a high penetration of renewable energies. In order to train ANNs appropriately with reliable estimation accuracy, the scenario generator was presented, which captures a feasible number of grid states including the volatile power generation. In addition, it was shown how a selection of hyperparameters can be achieved with a reasonable trade-off between estimation accuracy and training time. Furthermore, WLS SE and ANN were compared with a variety of simulations using a benchmark grid (CIGRE test grid) [Ru06] and a real German distribution system. This involved the analysis and evaluation of various test cases with different measurement configurations, topological errors and bad data. In all generated results of the test cases, the ANN monitoring method outperformed the WLS SE. The methods presented in this paper extend the previous approaches by enabling the training of ANNs through historical feed-in time series of individual generations in order to improve the estimation accuracy of the ANNs.

3 Artificial Neural Networks for Active and Reactive Power Monitoring Of Medium-Voltage Grids

3.1 ANN Fundamentals

A fully connected feed-forward ANN consists of individual components called neurons. The model of a neuron can receive different input values $X_1, X_2, X_3, \ldots X_n$, which are weighted with the respective weights $w_{1j}, w_{2j}, w_{3j}, \ldots w_{nj}$ and then summed up. The weighted and summed input values then pass an activation function φ. This can be represented, for example, by one of the following functions:

$$\varphi_{\text{ReLU}}(x) = \max(0, x) \qquad \varphi_{\text{Sigmoid}}(x) = \frac{1}{1 + \exp(-x)} \qquad \varphi_{\text{Tanh}}(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$$

A single neuron outputs Y as the output value. The flow of information through such a neuron is shown in Fig. 2. Typically, several parallel neurons are combined as so called

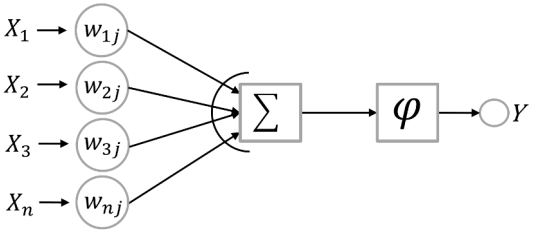


Fig. 2: Scheme of a neuron

layers. Serially connected layers form the complete artificial neuronal network. All outputs of the neurons of a previous layer are connected to every single neuron of the next layer.

Fig. 3 illustrates an exemplary fully connected feed-forward ANN with the input values $X_1, X_2, \ldots X_n$ and the output values $Y_1, Y_2, \ldots Y_n$. The training of the ANN runs over several epochs, in which an optimization algorithm (e.g. ADAM [KB14]) adjusts the individual weights $w_{1j}, w_{2j}, w_{3j}, \ldots w_{nj}$ of the ANN in such a way that they match the required outputs

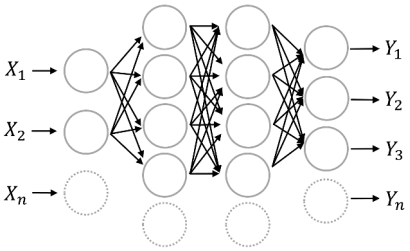


Fig. 3: Feed-forward ANN

and minimize a loss function. The training and testing process requires two different data sets. The so-called training set is applied for the training of the ANN, whereas the test set is used for the validation of the predictions made by the ANN. For a more detailed insight into the modelling of artificial neuronal networks, see [BL17] [SFM12] [Pa17].

3.2 Monitoring Electrical Distribution Grid Parameters with ANN

The static state of the electrical grid can be described by the power flow equations and can be determined by performing iterative power flow calculations. In distribution grids, the problem lies in the availability of necessary initialization variables due to the very low density of measuring devices. ANN can be applied here to describe the output parameters as approximate functions even with a small number of input values. By suitable training, the ANN can establish a link between the input parameters and the corresponding estimates of the output values. Fig. 4 shows the approach in which a trained ANN uses a small number of measurement values and the respective switching configuration to estimate the active power P, the reactive power Q, and the voltage magnitude V for each bus.

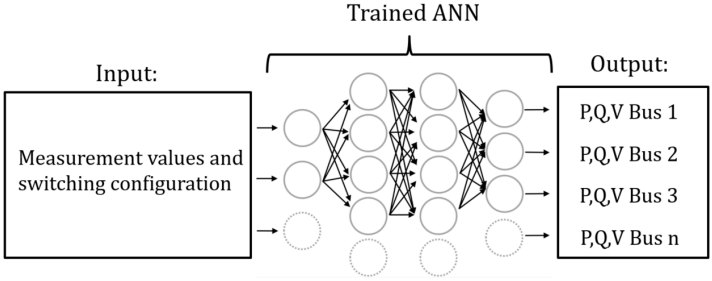


Fig. 4: ANN for monitoring electrical distribution grid parameters

4 ANN Training with Time Series

4.1 Generating Feed-in Time Series

The generation of feed-in time series is enabled via a Python interface to the platform renewables.ninja [PS16], which can run simulations of hourly power output from PV systems and wind power plants with a specific location [SP16]. The parametrization of the PV and wind models can be adjusted individually (e.g. turbine model, hub height, capacity, etc.). Furthermore, these feed-in models are based on weather data generated from satellite observations and hourly reanalysis [Pf19] [GM17]. Since the SimBench medium-voltage grid used for the validation is available in pandapower [Th18], we can create an individual feed-in time series based on the geo-coordinates for each PV system and wind power plant directly via the Python interface. In the model generation of PV systems, a distinction can be made between two weather data sets. For our simulations, we use the "Surface Solar Radiation Data Set - Heliosat" (SARAH), which provides hourly averages on a regular latitude/longitude with a spatial resolution of $0.05° \times 0.05°$ [Pf19].
Fig. 5 illustrates the corresponding feed-in time series over selected summer days in the years 2012 to 2015 for a 1.5 MW PV system connected to the medium-voltage level. Fig. 6 shows a feed-in time series for a wind power plant with a nominal power of 2.4 MW covering the first days of January in the years 2012 to 2015.

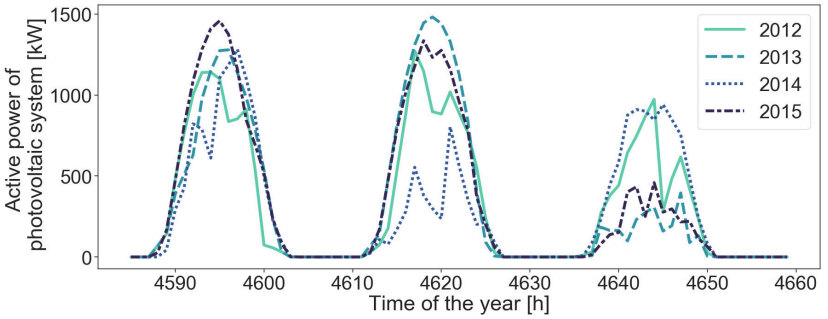


Fig. 5: Exemplified hourly feed-in time series for a 1.5 MW PV System

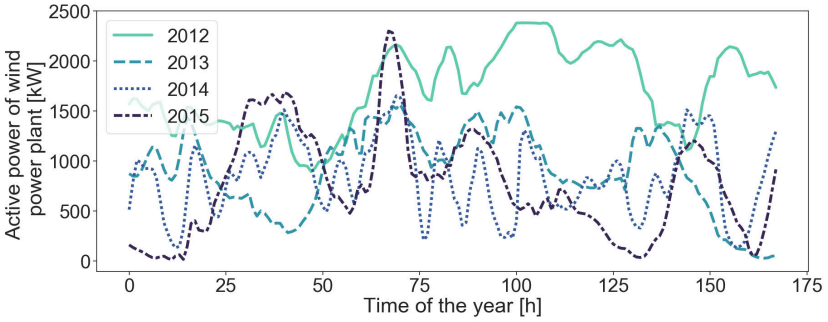


Fig. 6: Exemplified hourly feed-in time series for a 2.4 MW wind power plant

4.2 Load Time Series

For the loads, standard load profiles for industrial and household consumers are selected [BD19]. Since a high number of connected consumers are aggregated from the low-voltage level, an approximation can be made by load time series. In addition to the feed-in time series, the hourly load values of the profiles are scaled accordingly for each time step. To address the variability between individual loads, Gaussian noise is applied with a standard deviation of 5% for each load.

4.3 Generating the Training and Validation Data Set

To ensure that the ANNs are capable of estimating the relationship between the input measurements and the output values Q, P, V, and I reliably, it is essential to consider as many different grid states as computationally feasible during the training process.
First of all, the corresponding feed-in time series are stored for all individual PV and wind power plants of the electrical grid. In addition, the load time series are also saved. Subsequently, the components are scaled for every hourly time step based on the time series and a power flow calculation is performed. For each result of every time step, the active power P, reactive power Q, voltage magnitude V, and line current I for each bus and line are extracted and stored. Similarly, the corresponding measured values of the measuring equipment and the switch configuration are also saved. This procedure can be summarized as follows:

1. Generate and store feed-in time series and load time series
2. Scale the individual components for every time step accordingly
3. Perform a power flow calculation for each time step
4. Store the result of P, Q, V, I, the measurement values, and the switching configuration for each time step

The final training and test data set for the ANN therefore consists of the measured values and the switch configuration which serve as input parameters. The associated results of active power P, reactive power Q, voltage magnitude V, and line current I are set as target values.

5 Electrical Test Grid

The SimBench benchmark datasets [SBDfN19] were developed for a realistic analysis, planning, and management of distribution and transmission systems. For the simulation

of the estimation accuracy, the slightly modified SimBench 20 kV medium-voltage grid (semi-urban) is used, which was constructed on the basis of real German medium-voltage grids and features geographical allocation.
The grid operated as an open ring and is connected to the high-voltage level via two parallel transformers. A total of nine feeders exist. Table 1 summarises the sum of connected loads and static generators (PV generators and wind power plants) of each feeder. The grid consists of 117 buses, 121 lines, 115 loads, 116 PV generators and 5 wind power plants. Of the 117 buses, 4 are virtual, so they are electric sleeves on which no active or reactive power injection can occur. Moreover, three plausible switching configurations were selected; the first configuration is shown in Fig. 7. In the SimBench grid, measuring devices are located

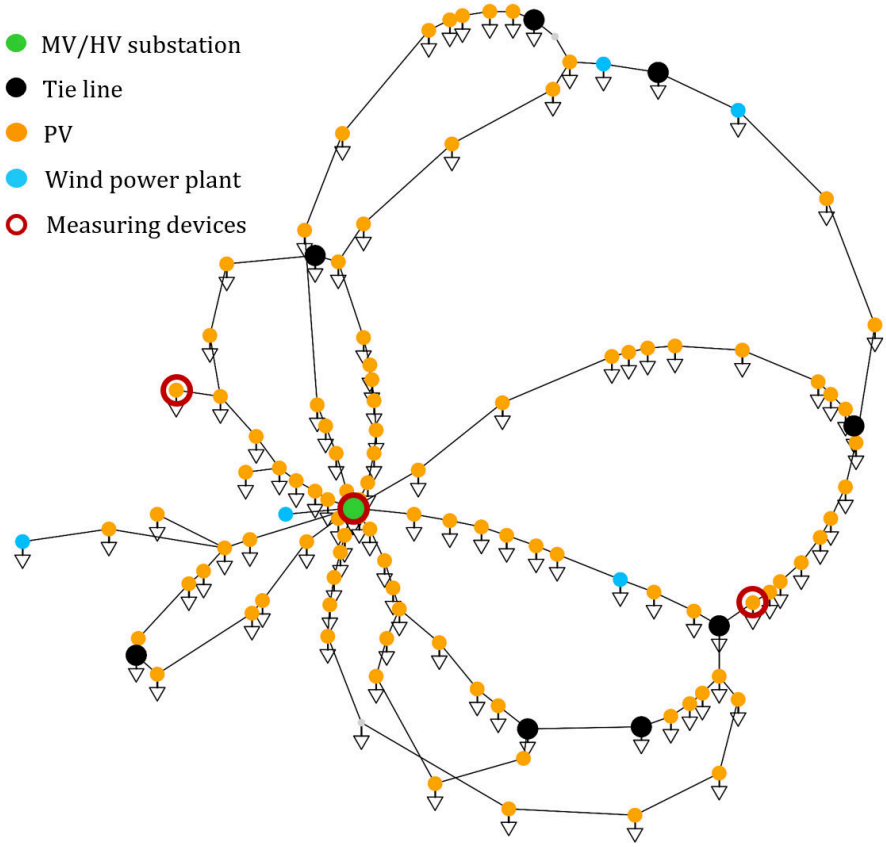


Fig. 7: SimBench medium-voltage grid with the default configuration of tie lines

at three buses. They measure active power P, reactive power Q, and voltage magnitude V. In addition, in the first line of each feeder outgoing from the substation, P and Q are also measured. Thus, a total of 27 measurements are available. Measurement device tolerance is

given by accuracy classes, which describe the accuracy that can be expected for a measured value. AC 0.5 is applied for voltage measurements. This class is described for instance in IEC 61869-2. The associated standard deviation is $\frac{0.5}{3} \approx 0.167\,\%$. Thus 99.7 % of all measurements have a maximum error of 0.5 % within the $3\,\sigma$ confidence interval. For measurements of P and Q, a standard deviation of $\frac{2}{3}$ % is assumed. This level of accuracy is considered for all power flows and is consequently present in the training and test sets for the ANN.

Feeder	Number of loads [n]	Total active power of loads [MW]	Number of static generators [n]	Total active power of static generators [MW]	Total line length [km]
1	22	7.427	24	5.987	11.89
2	12	3.016	12	1.182	7.03
3	11	2.603	12	2.189	5.60
4	12	2.881	14	1.292	6.73
5	9	1.383	10	4.546	6.25
6	6	1.596	6	0.828	3.40
7	11	2.976	12	1.512	6.87
8	16	5.496	16	3.287	7.26
9	15	3.922	14	1.471	8.34
substation	1	0.340	2	3.000	-
sum	115	31.64	122	25.29	63.37

Tab. 1: Feeder parameters of the SimBench 20 kV medium-voltage grid (semi-urban)

6 Simulation

The fully connected feed-forward ANN presented in chapter 3.1 was used for the simulations. The estimation accuracy could not be increased by using a recursive neural network (RNN) based on long short-term memory (LSTM). With the research focus on time series with higher time resolution and forecasting capability for monitoring with ANN, RNN (especially RNN with an attention mechanism) will be considered. Each of the parameters active power P, reactive power Q, voltage magnitude V, and line current I is trained on a single ANN over 500 epochs using the open source deep learning library PyTorch [Pa17]. The ANN models consist of three layers, with a hidden layer size of 500. The combination of hyperparameters was selected by the hyperband method [Li16] with the objective of maximum estimation accuracy.

The criteria C1 and C2 from [MBB19] are selected for the evaluation of the simulation results. C1 is fulfilled if the bus voltage magnitude errors are less than 1 % (C2: $U_\Delta < 0.5\,\%$) and line loading errors are less than 10 % (C2: $I_\Delta < 5\,\%$). For the simulation and validation process, three test sets and a training set based on the procedure outlined in Section 4.3 are generated. The training set consists of 10 years in total (2003-2012). This set is used to train the ANN. In addition, three plausible switching configurations of the SimBench grid are

Fig. 8: Percentage error of the total estimated reactive and active power of the SimBench grid (2015)

considered for each year. For the training set, this results in 263016 hourly time steps (87672 hours over all years for 3 switching configurations). The years 2013, 2014, and 2015 are individually used as test sets. Each of them consists of 26280 time steps (8760 hours for 3 switching configurations). Fig. 8 illustrates the percentage error of the total estimated active and reactive power of the SimBench grid for each time step in the test year 2015. In this case the mean error for monitoring the reactive power is 2.047 % (max: 4.031 %) and the active power is 2.037 % (max: 4.922 %). Fig. 9 shows the estimation errors of the reactive power as estimated by the ANN at each bus over the hourly time steps of the test year 2015,

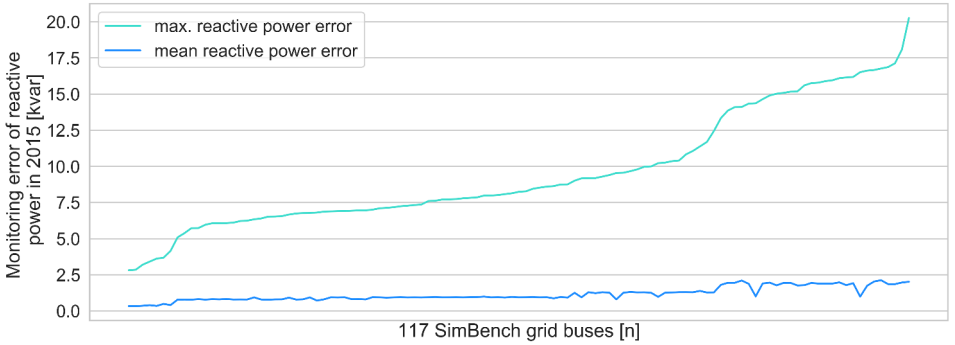


Fig. 9: Monitoring error of reactive power at the SimBench grid (2015)

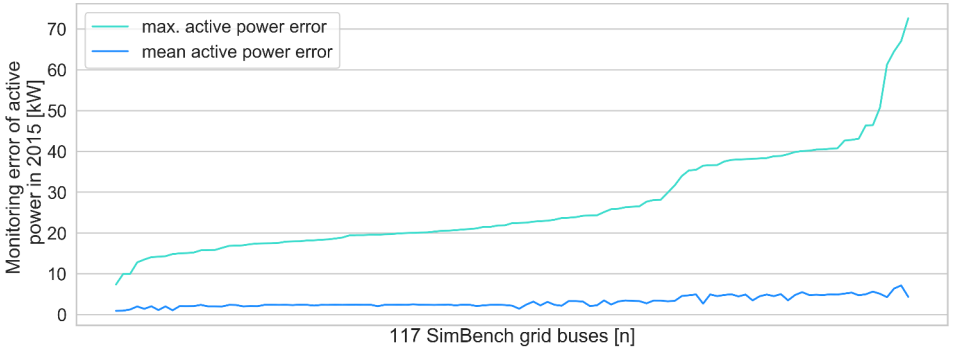


Fig. 10: Monitoring error of active power at the SimBench grid (2015)

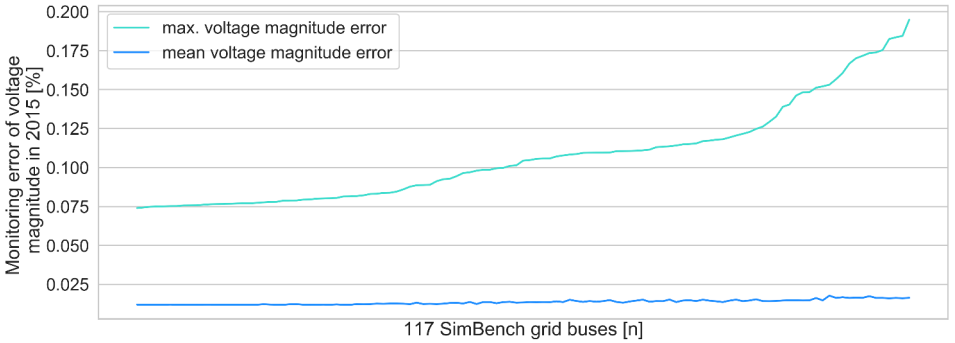


Fig. 11: Monitoring error of voltage magnitude at the SimBench grid (2015)

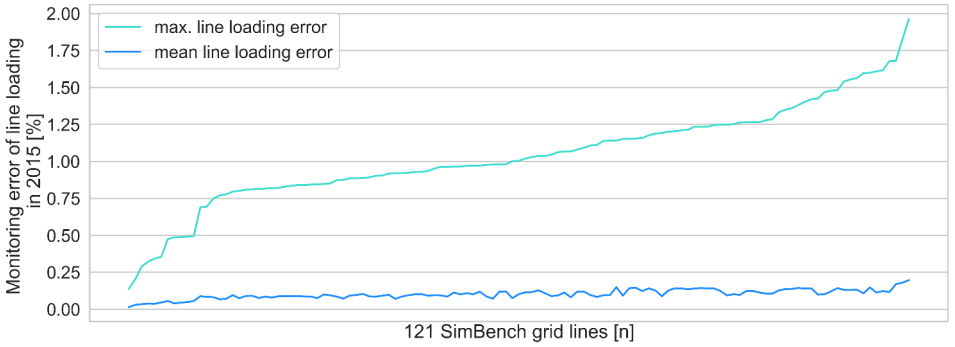


Fig. 12: Monitoring error of line loading at the SimBench grid (2015)

including the three switching configurations. Each individual boxplot indicates the under 0.25 quantile, the upper 0.75 quantile, the median, as well as the outliers. The maximum absolute error for estimating the reactive power over all time steps of the year 2015 is

20.254 kvar. The mean of the error lies at 1.258 kvar with a standard deviation of 1.342 kvar. The peak error for the active power estimation is 72.609 kW. In addition, the mean value of the error is 3.332 kW ± 3.620 kW (Fig. 10). Fig. 11 illustrates the maximum absolute estimation error for the voltage magnitude of 0.195 % and Table 2 shows the monitoring estimation errors for P and Q over the various test years. For the estimation errors of bus voltage magnitude and line loadings, the strict criterion C2 ($U_\Delta < 0.5$ %, $I_\Delta < 5$ %) was met in all simulation results. If the criteria of line loadings are set as the maximum monitoring error of the total reactive and active power in the grid, C2 is fulfilled for the reactive power estimation in all test years, whereas criterion C1 can be achieved for the monitoring of active power for the years 2013 and 2014.

		Maximum Absolute Error	Mean Absolute Error (MAE)	Standard Deviation (SD)	Maximum percentage error of the total active and reactive power in the grid [%]
Test year 2013	P	60.272 kW	3.334 kW	±3.621 kW	5.084 %
	Q	18.246 kvar	1.257 kvar	±1.343 kvar	3.891 %
Test year 2014	P	62.902 kW	3.331 kW	±3.622 kW	5.175 %
	Q	20.251 kvar	1.261 kvar	±1.347 kvar	4.296 %
Test year 2015	P	72.609 kW	3.332 kW	±3.620 kW	4.922 %
	Q	20.254 kvar	1.258 kvar	±1.342 kvar	4.031 %

Tab. 2: Overview of monitoring errors for each test year

In the following, the approach of training with time series is compared with the scenario generator, which is presented in [MBB19]. The scenarios generated by the scenario generator are referred to as scaled scenarios. In order to establish comparability in the number of training scenarios, both training sets consist of approximately 260000 training samples. Again, the active power in the target year 2015 is estimated. The mean absolute error and

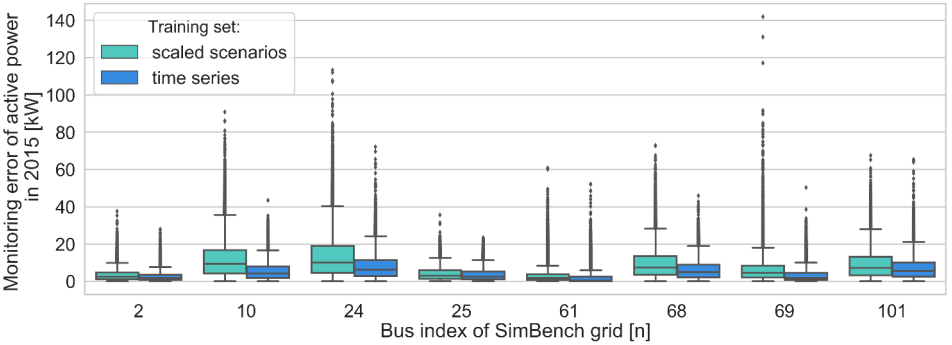


Fig. 13: Monitoring error of active power at the SimBench grid (2015)

the standard deviation for the estimation of active power are 3.818 kW ± 4.384 kW using the scaled scenarios (time series: 3.332 kW ± 3.620 kW). The difference in mean and standard deviations are mainly explained by the estimation of active power at PV and wind power

plants. Fig. 13 shows the difference in the estimation error among the 8 largest PV and wind power plants in the SimBench grid. As the ANN has learned the individual feed-ins over a period of 10 years, the estimation error is lower for the ANN trained by time series.

7 Conclusion and Outlook

This paper presents an approach for monitoring active power P, reactive power Q, voltage magnitude V, and line current I in medium-voltage grids based on ANN. For the simulation and evaluation process, the estimation accuracy of the ANN was demonstrated using three separate test years (2013, 2014, and 2015). The errors for estimating V (0.195 %) and I (1.962 %) are sufficient to identify current and voltage violation with high accuracy. In addition, the mean errors for estimating the active (2.037 %) and reactive power (2.047 %) over the entire grid are sufficiently low to provide essential information for optimization strategies at higher voltage levels which has been verified by meeting criteria C1/C2. Conclusively, the accuracy of the estimation increased especially on large PV systems and wind power plants when time series are used for training the ANN. This demonstrated that the individual feed-in characteristics influenced by weather phenomena can be learned from the ANN and thus reduce the estimation error.
However, there are also some drawbacks of the presented method. The behavior of individual loads can be significantly more diverse than assumed and may differ from standard load profiles. In addition, only time series with an hourly interval was used for the validation. A higher time resolution may be considered for further evaluations. Limitations are primarily related to dynamic topologies, e.g. many switching states or transformers with taps, which significantly extends the training time. Furthermore, future research could evaluate if constant retraining of the ANN with new time series data could further decrease the estimation errors. In this way, it could be ensured that further grid states continually update the ANN.

Bibliography

[AGE04] Abur, A; Gomez-Exposito, Antonio: Power System State Estimation: Theory and Implementation, volume 24. 01 2004.

[BD19] BDEW Bundesverband der Energie- und Wasserwirtschaft e.V.: , 2019. (www.bdew.de, accessed 22.04.2019).

[BL17] Buduma, Nikhil; Locascio, Nicholas: Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms. O'Reilly Media, Inc., 1st edition, 2017.

[Bu19] Burger, B.: , Energy charts - Fraunhofer ise, 2019.

[DEA13] Deutsche-Energie-Agentur: , dena Verteilnetzstudie. Ausbau- und Innovationsbedarf der Stromverteilnetze in Deutschland bis 2030, December 2013.

[GM17] Gelaro, Ronald; McCarty, Will et al.: The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). Journal of Climate, 30(14):5419–5454, 2017.

[IG12] Ivanov, O.; Garvrilaş, M.: State estimation for power systems with multilayer perceptron neural networks. In: 11th Symposium on Neural Network Applications in Electrical Engineering. pp. 243–246, Sep. 2012.

[KB14] Kingma, Diederik; Ba, Jimmy: Adam: A Method for Stochastic Optimization. International Conference on Learning Representations, 12 2014.

[Li16] Li, Lisha; Jamieson, Kevin G.; DeSalvo, Giulia; Rostamizadeh, Afshin; Talwalkar, Ameet: Efficient Hyperparameter Optimization and Infinitely Many Armed Bandits. CoRR, abs/1603.06560, 2016.

[MBB19] Menke, Jan-Hendrik; Bornhorst, Nils; Braun, Martin: Distribution System Monitoring for Smart Power Grids with Distributed Generation Using Artificial Neural Networks. International Journal of Electrical Power and Energy, 113:472–480, December 2019.

[OM14] Onwuachumba, A.; Musavi, M.: New Reduced Model approach for Power System State Estimation Using Artificial Neural Networks and Principal Component Analysis. In: 2014 IEEE Electrical Power and Energy Conference. pp. 15–20, Nov 2014.

[OWM13] Onwuachumba, A.; Wu, Y.; Musavi, M.: Reduced Model for Power System State Estimation Using Artificial Neural Networks. In: 2013 IEEE Green Technologies Conference (GreenTech). pp. 407–413, April 2013.

[Pa17] Paszke, Adam; Gross, Sam; Chintala, Soumith; Chanan, Gregory; Yang, Edward; DeVito, Zachary; Lin, Zeming; Desmaison, Alban; Antiga, Luca; Lerer, Adam: Automatic differentiation in PyTorch. In: NIPS-W. 2017.

[Pf19] Pfeifroth, Uwe; Kothe, Steffen; Trentmann Jörg; Hollmann Rainer; Fuchs Petra; Kaiser Johannes; Werscheck Martin: Surface Radiation Data Set - Heliosat (SARAH) - Edition 2.1. Satellite Application Facility on Climate Monitoring (CM SAF), 2019.

[PS16] Pfenninger, Stefan; Staffell, Iain: Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy, 114:1251 – 1265, 2016.

[Ru06] Rudion, K.; Orths, A.; Styczynski, Z. A.; Strunz, K.: Design of benchmark of medium voltage distribution network for investigation of DG integration. In: 2006 IEEE Power Engineering Society General Meeting. June 2006.

[SBDfN19] SimBench-Benchmark-Datensatz für Netzanalyse, Netzplanung und Netzbetriebsführung: , 2019. (www.simbench.de, accessed 21.04.2019).

[SFM12] Shiffman, D.; Fry, S.; Marsh, Z.: The Nature of Code. 2012.

[SP16] Staffell, Iain; Pfenninger, Stefan: Using bias-corrected reanalysis to simulate current and future wind power output. Energy, 114:1224 – 1239, 2016.

[Th18] Thurner, L.; Scheidler, A.; Schäfer, F.; Menke, J.; Dollichon, J.; Meier, F.; Meinecke, S.; Braun, M.: pandapower — An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems. IEEE Transactions on Power Systems, 33(6):6510–6521, Nov 2018.

Der CO_2-Kompass: Konzeption und Entwicklung eines Tools zur emissionsarmen Stromnutzung

Lucas Hüer[1], Nico Stadie[1], Simon Hagen[2], Oliver Thomas[2], Hans-Jürgen Pfisterer[1]

Abstract: Um Elektromobilität nachhaltiger zu gestalten, muss es die Möglichkeit geben, Elektrofahrzeuge zu jenen Zeitpunkten zu laden, an denen der Strom zu einem großen Teil aus erneuerbaren Quellen generiert wird. Hierfür wurde mit Hilfe der Scrum-Methode ein Software-System entwickelt, welches Endkunden die aktuelle Zusammensetzung des Strommix transparent anzeigen kann: Der CO_2-Kompass. In diesem Beitrag wird die Entwicklung des CO_2-Kompass vorgestellt. Zudem soll verdeutlicht werden, warum dieses Tool wichtig für eine nachhaltigere, emissionsarme Stromnutzung in der Elektromobilität sein kann. Dabei wird nicht nur auf die Notwendigkeit des Systems als Dienstleistung eingegangen, sondern es wird auch beschrieben wie das System aufgebaut ist und wie es in ein bestehendes Produkt (in diesem Fall eine Ladesäule) integriert werden kann.

Keywords: CO_2-Kompass, Elektromobilität, Nachhaltigkeit, CO_2-Emissionen, Ladesäule, Energieversorgung

1 Einleitung

Das noch junge 21. Jahrhundert wird bislang zu einem großen Teil von Erkenntnissen, Diskussionen und Entscheidungen rund um den Klimawandel geprägt [Br17]. Regierungen der verschiedensten Länder kommen regelmäßig zusammen und einigen sich auf diverse Abkommen, um Umwelt und Gesellschaft vor den negativen Folgen des Klimawandels zu schützen. Einer der größten Faktoren der in den bisherigen Klimagipfeln angesprochen wurde, ist der Ausstoß von Treibhausgasen [Pr15]. Um die Emission dieser Gase einzudämmen und damit auch den Treibhauseffekt zu bekämpfen, der primär zur globalen Erwärmung führt, setzt sich die Weltgemeinschaft immer ehrgeizigere Ziele. Für Deutschland hat die Bundesregierung beschlossen, die Treibhausgasemission im Zeitraum von 1990 bis 2020 um 40% zu reduzieren [SE13]. Für die darauffolgenden Jahrzehnte sind die Ziele noch ehrgeiziger gesteckt. Daher ist es notwendig innovative Produkte und Dienstleistungen zu entwickeln, die Unternehmen und Verbraucher bei der Einsparung von CO_2-Emissionen unterstützen. Ein besonders hohes Potenzial zur Einsparung liegt im Mobilitäts-Sektor, was auch durch die Bundesregierung erkannt wurde. Daher plant die Nationale Plattform Elektromobilität (NPE) den heimischen Markt als einen Leitmarkt für Elektromobilität aufzustellen und bis 2020 einen Bestand von einer Million elektrischen Fahrzeugen zu erreichen [NA14]. Diese verursachen bei Fahrten

[1] Hochschule Osnabrück, Ingenieurswissenschaften und Informatik, Albrechtstr. 30, 49076 Osnabrück, l.hueer@hs-osnabrueck.de
[2] Universität Osnabrück, Informationsmanagement und Wirtschaftsinformatik, Katharinenstraße 3, 49074, simon.hagen@uni-osnabrueck.de

doi:10.18420/inf2019_72

keinerlei CO_2-Ausstoß, was jedoch nicht davon ablenken darf, dass der Strom zum Laden der elektrischen Fahrzeuge oftmals nicht CO_2-frei generiert wird [Sc15]. Der derzeitige Strommix in Deutschland besteht sowohl aus konventionellen Energiequellen (z.B. Braunkohle, Erdöl oder Kernenergie) wie auch aus erneuerbaren Energiequellen (z.B. Windkraft, Wasserkraft oder Photovoltaik) und obwohl der Anteil an erneuerbaren Energien stetig steigt, betrug der Anteil der konventionellen Energien im Jahr 2018 noch 64,8%. [Sc19; basierend auf AGEB 2018] Da es laut Trommer et al. den meisten Leuten allerdings wichtig ist, Elektrofahrzeuge mit Strom aus erneuerbaren Energiequellen aufzuladen [Tr13], müssen innovative Lösungen angeboten werden, die es dem Kunden gestatten, einen transparenten Überblick über den Strommix zu bekommen. Ziel dieses Beitrags ist es daher, mit Hilfe einer web-basierten Plattform, dem CO_2-Kompass, den Endkunden die entsprechende Transparenz zu ermöglichen und eine Einbindung in automatische Ladesysteme vorzubereiten. Da ein solches System bisher nicht existiert, wurde nach der Erhebung von Anforderungen in einem Scrum-Projekt das Artefakt entwickelt und im Folgenden detailliert vorgestellt. Um das Ziel des Beitrages zu erreichen, ist dieser wie folgt strukturiert: Nachdem die angewandten wissenschaftlichen Methoden in Kapitel 2 erläutert werden, wird dem Leser in Kapitel 3 das nötige Hintergrundwissen vermittelt indem existierende Literatur analysiert wird. Dadurch soll vor allem verdeutlicht werden, wieso die Erstellung des CO_2-Kompasses wichtig ist. Hierfür wurde relevante deutsche und englische Literatur herangezogen. Im Anschluss wird dem Leser in Kapitel 4 beschrieben, wie der CO_2-Kompass funktioniert und welche Muss- und Wunschkriterien bei der Erstellung des Systems berücksichtigt wurden. Im Anschluss an die Projektbeschreibung, wird in Kapitel 5 dargestellt, wie der CO_2-Kompass angewendet werden kann. Hier wird besonders auf eine mögliche Integration der Software in Ladesäulen eingegangen. Abschließend werden die Ergebnisse dieses Artikels zusammengefasst und diskutiert.

2 Methodik der Arbeit

Zur Entwicklung des CO_2-Kompass wurde der Scrum-Ansatz [SC97] gewählt und in diesem Kontext ein „Product Backlog“ erstellt, welches die funktionellen und nichtfunktionellen Systemanforderungen beinhaltet. Um diese Anforderungen aufzustellen, wurde ein Interview mit einem Experten aus dem Bereich der Energiewirtschaft und des Ladesäulenmanagements durchgeführt. Aus den daraus resultierenden Anforderungen wurden die Arbeitspakete für die Sprint-Phasen abgeleitet und dem Entwicklerteam fortlaufend zugewiesen. Während der Entwicklungsphase wurden die Arbeitspakete unter Aufsicht des Projektleiters bearbeitet und in „Daily Scrum´s“ beurteilt und nach Vervollständigung als erledigt markiert. Nach Abschluss jeder Sprintphase wurde das Werkzeug evaluiert und das nächste Arbeitspaket basierend auf den Ergebnissen begonnen. So entstand inkrementell das den Anforderungen entsprechende Tool des CO_2-Kompass. Die agile Softwareentwicklungsmethode Scrum wurde ausgewählt, um die Produktivität der Teamarbeit durch gemeinsame Zielsetzung, Definition von Teilaufgaben und Transpa-

renz des Projektfortschritts zu verbessern [GL10], was sich im vorliegenden Projekt als positiv herausgestellt hat.

3 CO_2-Emissionen in der Elektromobilität

Trotz ehrgeiziger Ziele der Bundesregierung, die nach dem Pariser Klimaabkommen „eine praktisch vollständige Reduktion der CO_2-Emissionen des Verkehrs“ [Pl18, S. 2] bis 2050 vorsieht, hat sich der verkehrsbezogene CO_2-Ausstoß in den letzten Jahren vervielfacht [Zi16]. Daher wird es immer wichtiger Lösungen zu finden, die den Güter- und Personenverkehr emissionsfrei gestalten können. Hierbei ist nicht nur die Vermeidung des Verkehrs eine Option, sondern auch der Wechsel zu alternativen Verkehrsmitteln und Antrieben. Während der Güterverkehr von Lastkraftwagen nicht so einfach auf Züge oder ähnliche Alternativen mit CO_2-Einsparpotenzial verlagert werden kann, gibt es für den Personenverkehr eine technische Alternative, die von vielen Experten als durchführbar und plausibel angesehen wird: Der Umstieg auf batterieelektrische Fahrzeuge [Pl18]. Laut dem Umweltbundesamt [siehe Ze19] trägt der Verkehr mit etwa 19 Prozent zum gesamten CO_2-Ausstoß in Deutschland bei, wobei die Emissionen in den Jahren seit 2010 trotz Innovationen in der Antriebstechnik stark gestiegen sind. Dieser zusätzliche CO_2-Ausstoß kann auf die größere Anzahl an Fahrzeugen mit Verbrennungsmotoren (Benzin und Diesel) zurückgeführt werden [Ze19]. Die Elektromobilität spiegelt einen Lösungsansatz wieder, der die ansteigenden Emissionen der Verbrenner-Autos verdrängen könnte; denn durch den elektrischen Antrieb verursachen Elektroautos keinen direkten Stickoxid- und CO_2-Ausstoß sowie weniger Feinstaub-Emission während der Fahrtzeiten. Es muss allerdings beachtet werden, dass die Produktion des Ladestroms für erhebliche indirekte CO_2-Emissionen sorgen kann; daher sollte eine Gesamtbetrachtung (Well-to-Wheel Ansatz) der Emissionen erfolgen [Ha11]. Im Gegensatz zum Tank-to-Wheel Ansatz, welcher nur jene Energie betrachtet, die von einem Fahrzeug aufgenommen und umgewandelt wird, kann mit dem Well-to-Wheel Ansatz die gesamte Wirkungskette betrachtet werden. Es wird also auch die Energiebereitstellung in die Betrachtung eingeschlossen [Th14; Or16]. In Deutschland wird die Energie über verschiedene erneuerbare und konventionelle Energiequellen hergestellt. Zu den Primärenergieträgern gehören neben Kernenergie, Erdgas, Steinkohle, Pumpspeicherwasser, Braunkohle und sonstige konventionelle Quellen auch folgende erneuerbare Energieträger: Photovoltaik, Geothermie, Windkraft, Laufwasser / Speicherwasser mit natürlichem Zufluss und Biomasse [Um17].

Indem erneuerbare Energien die fossilen Energieträger im deutschen Strommix ersetzen, können Treibhausgase vermieden werden. Dies spiegelt sich im Jahr 2018 in der Vermeidung von 183,7 Millionen Tonnen CO_2-Äquivalente (Einheit zur vereinheitlichten Messung unterschiedlicher Treibhausgase) für Deutschland wieder, was einer Verdopplung der Treibhausgas-Vermeidung gegenüber 2010 entspricht [Um19]. Es gibt also zunehmend Alternativen zur CO_2-Emission, die durch konventionelle Energiequellen ermöglicht wird. Zudem gibt es viele Einwohner in Deutschland, die ihre eigene CO_2-

Bilanz verbessern möchten. Dies spiegelt sich zum Beispiel in einer Befragung wieder, in der die Teilnehmer zu großen Teilen (über 80%) angaben, dass der Ladestrom für Elektromobilität aus erneuerbaren Quellen generiert werden soll [Tr13]. Da es also möglich ist Strom emissionsfrei herzustellen und viele Endkunden Strom aus nachhaltigen Energiequellen bevorzugen, wird eine transparente Darstellung des Strommixes immer wichtiger um dem umweltbewussten, mündigen Endkunden einen eigenen Entscheidungsspielraum zu geben.

4 Der CO_2-Kompass

Um die Vorteile regenerativer Energien stärker nutzen zu können, ist es hilfreich, die Stromnutzung von elektrischen Verbrauchern zeitlich an die Energiegewinnung anzupassen. Dadurch können elektrische Verbraucher zu jenen Zeitpunkten genutzt werden, an denen der regenerative Anteil der Stromgewinnung besonders hoch ist. Hierfür wurde die Software des CO_2-Kompass entwickelt, welche die Stromerzeugung und die einhergehende CO_2-Emissionen aufzeichnet und zudem in der Lage ist, Prognosen über zukünftige Zusammensetzungen des Strommixes zu geben. Diese Informationen lassen sich mit der Nutzung von energieintensiver Hardware wie zum Beispiel Wärmepumpen, Klimaanlagen oder Produktionsmaschinen koppeln. In einem ersten Schritt sollen die Daten allerdings genutzt werden, um Ladevorgänge für Elektrofahrzeuge systematisch mit erneuerbarer Energie zu versorgen. Es kann somit entschieden werden, ob es zu einem bestimmten Zeitpunkt ökologisch sinnvoll ist, sein Elektrofahrzeug aufzuladen (siehe Kapitel 5.2). Um dies zu gewährleisten, werden die Produktionsdaten der einzelnen Energieversorger von der ENTSO-E-Datenbank[3] (Verband Europäischer Übertragungsnetzbetreiber) abgegriffen und zur weiteren Verarbeitung auf einem zentralen Server gespeichert. Diese Daten werden detailliert ausgewertet und ein Algorithmus sorgt für die Ermittlung der CO_2-Emissionen aller Produktionsarten (Solarkraft, Windkraft, Atomkraft etc.). Bevor in 4.2 beschrieben wird, wie der CO_2-Kompass aufgebaut ist, werden im folgenden Unterkapitel die Anforderungen an das System aufgelistet.

4.1 Anforderungen

Bevor der CO_2-Kompass entwickelt und umgesetzt werden konnte, wurde ein Anforderungsprofil im Zuge eines Experteninterviews mit einem Fachmann aus dem Bereich Energiewirtschaft und Ladesäulenmanagement aufgestellt, welches sechs Muss- und drei Wunschkriterien beinhaltet. Zu den Musskriterien gehören all die Aspekte, welche unabdingbar für den CO_2-Kompass sind:

- Kontinuierlicher Bezug der Daten über die aktuelle Stromproduktion aus der ENTSO-E Transparency Datenbank. Unter Daten ist in diesem Kontext die Information über die Menge des erzeugten Stromes in Abhängigkeit zu der Erzeu-

[3] https://www.entsoe.eu/data/

gungsart, gruppiert nach den regionalen Netzbetreibern, sowie die Gesamtheit aller Netzbetreiber in Deutschland, zu verstehen.

- Kontinuierlicher Download der Daten über die voraussichtliche zukünftige Stromproduktion für den nächsten Tag aus der ENTSO-E Transparency Datenbank.
- Persistente und konsistente Speicherung der erhaltenen und berechneten Daten in einer eigens verwalteten Datenbank.
- Laufende Berechnung der aktuellen CO_2-Emissionen (gemessen in gCO_2eq/kWh) auf Grundlage der abgerufenen Daten.
- Kontinuierliche Berechnung des CO_2-Ausstoßes bei der Stromproduktion für den Folgetag durch Verwendung der abgespeicherten Daten.
- Bereitstellung einer Software-Schnittstelle, die es Klienten ermöglicht, die bereitgestellten Daten des CO_2-Kompass für Entscheidungen zum Ladevorgang eines Elektroautos zu nutzen.

Die drei Wunschkriterien sind zwar - im Gegensatz zu den Musskriterien - nicht unabdingbar für das System, wurden aber im Zuge des Experteninterviews ausdrücklich gefordert. Die folgenden Kriterien entscheiden also nicht über den Betrieb des Systems, können allerdings die gewünschten Eigenschaften, welche aus dem Interview hervorgegangen sind, bereitstellen:

- Über eine Schnittstelle sollten dem Klienten die voraussichtlichen CO_2-Daten (ausgerechnet auf Basis der von ENTSO-E bereitgestellten Rohdaten) zur Verfügung gestellt werden.
- Der Zugriff auf die Schnittstelle sollte nach erfolgreicher Benutzerverifikation erfolgen können.
- In regelmäßigen Abständen soll eine Datensicherung der Datenbank angefertigt werden.

Der Anforderungskatalog, bestehend aus Muss- und Wunschkriterien, bildet die Grundlage für die Entwicklung der Software. Nach der Scrum-Methode wurden die Anforderungen nach für nach in das System gefügt, sodass eine funktionierende Architektur für den CO_2-Kompass aufgestellt wurde. Neben den beschriebenen Kriterien wurde auch ein weiteres Abgrenzungskriterium aufgestellt um dem Projekt einen realistischen Rahmen zu geben. Dieses Kriterium wird als Grundlage für zukünftige Forschungsfelder in Kapitel 6 erklärt.

4.2 Architektur

Durch einhalten und umsetzen der Kriterien wurde der CO_2-Kompass im Laufe eines mehrwöchigen Projektes entwickelt und folgt der in Abbildung 1 dargestellten System-Architektur. Der CO_2-Kompass ist in drei Teilsysteme unterteilt:

(1) Der Crawler: Eine Schnittstelle zwischen ENTSO-E und eigener Datenbank
(2) Der CO_2-Calculator: Schnittstelle zwischen Rohdaten und Vorhersagedaten
(3) Die REST Schnittstelle: Erlaubt öffentlichen Zugang zu den Daten

Diese Teilsysteme werden im Folgenden erläutert, um die Funktionen des CO_2-Kompass und das Zusammenspiel zwischen ENTSO-E, Datenbank und Webseite (bzw. Ladesäule) darzustellen.

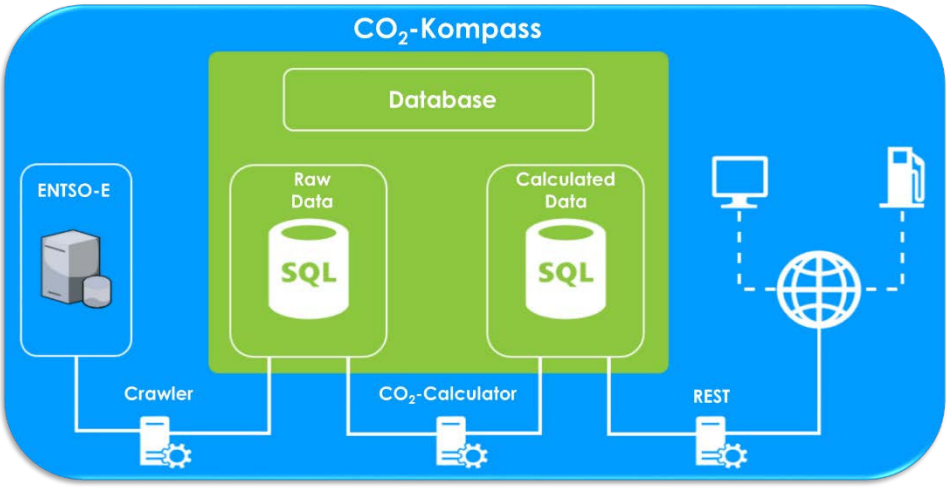


Abb. 1 - Architektur des CO_2-Kompass

Die benötigten Daten zur Stromproduktion in Deutschland werden durch ENTSO-E über eine Rest-API zur freien Verfügung bereitgestellt. Dies geschieht über einen angeforderten Sicherheitstoken. (1) Im Rahmen des Projektes werden diese Rohdaten von einem selbstentwickelten Crawler in fünfminütigen Abständen abgegriffen und auf Aktualisierungen geprüft. Fand eine Aktualisierung statt, so werden die Rohdaten strukturiert über ein Open-Source-Datenbankmanagementsystem in einer Datenbank abgelegt. Abgespeichert werden dabei die Stromproduktions-Daten für jeden Netzbetreiber in Deutschland und für Deutschland im Gesamten. Die Menge der derzeit 18 verschiedenen Produktionsarten des produzierten Stromes wird in Megawatt (MW) angegeben. (2) Weiterhin erfolgt alle fünf Minuten eine Berechnung der spezifischen CO_2-Werte für die vom Crawler hinzugefügten Produktionszahlen. Diese vom CO_2-Calculator berechneten Werte werden ebenfalls in der Datenbank abgespeichert und können so dem Netzbetreiber und der Produktionsart zugeordnet werden. Für die Berechnung der CO_2-Werte werden für jeden Netzbetreiber die Produkte von Produktionsmenge und spezifischer Emission pro Stromart (siehe [Sc14] für spezifische Emission) aufsummiert und zum Schluss durch die Summe an produziertem Strom geteilt. Diese Berechnung findet für alle fünf Anbieter in fünfminütigen Intervallen statt und liefert Werte im 15 Minuten Takt. Des

Weiteren erstellt der CO_2-Calculator jeden Tag um 0 Uhr eine CO_2-Vorhersage pro Produktionsart für den nächsten Tag. Dies geschieht für jeden Übertragungsnetzanbieter in Deutschland und die Vorhersage-Daten werden in der Datenbank abgelegt. Für die Vorhersage muss zunächst geprüft werden, ob die vom Netzbetreiber prognostizierten Produktionswerte für Wind und Solar vorliegen. Ist dies der Fall, wird weiterhin erst einmal geschaut, ob die Berechnung für den Betreiber bereits durchgeführt wurde. Ist dies nicht der Fall, startet die Berechnung. Folgende Formel wird genutzt, um den spezifischen Forecast nach Übertragungsenergieversorgern zu berechnen:

$$\frac{\sum_{k=0}^{Anzahl\ Produktionstypen} \frac{Produktionsmenge[k]\ von\ 3\ Tagen}{3} * Faktor * Spezifischer\ Emissionsausstoß}{\sum_{k=0}^{Anzahl\ Produktionstypen} \frac{Produktionsmenge[k] von\ 3\ Tagen}{3} * Faktor * Produzierte\ Menge\ in\ MW}$$

Dabei wird zunächst die Medianproduktion von allen Produktionsarten der letzten drei Tage berechnet, deren Produktion relativ konstant ist (z.B. Atomkraft, Öl). Diese Medianwerte werden mit ihren spezifischen CO_2-Ausstößen [siehe Sc14 für spezifische Emission] multipliziert und anschließend mit den Wind- und Solarvorhersagewerten, welche ebenfalls mit ihren spezifischen Emissionen multipliziert werden, aufaddiert. Dies geschieht immer jeweils für die gleichen Uhrzeiten. Danach werden die Strommengen der bereits berechneten Produktionsarten von dem vorhergesagten Gesamtstrom abgezogen. Man kann diesen hier auch als Reststrom bezeichnen. Dies sorgt für ein relativ stabiles Verhältnis aus vorhergesagten und bereits errechneten Werten. Der so übergebliebene Wert kann nun auf die übrigen Produktionsarten verteilt werden. Dafür werden für jede Produktionsart die Anteile an der Gesamtproduktion der letzten 2 – 4 Tage errechnet, so dass man einen Faktor erhält, der anschließend mit dem Reststrom multipliziert wird. Dies ergibt die Menge der voraussichtlichen Produktion für alle anderen Produktionsarten außer Wind und Solar. Die so errechneten Werte, werden durch den CO_2-Calculator berechnet und in die Datenbank eingetragen. (3) Um einen öffentlichen Zugang zu den Werten zu schaffen, wird eine REST Schnittstelle zur Verfügung gestellt. Diese erlaubt durch gezielte http-Anfragen einen Zugang zu den abgelegten Daten. Diese Schnittstelle kann weltweit von allen Clients abgegriffen werden.

5 Anwendung des CO_2-Kompass

Die Einsatzmöglichkeiten für die Verwendung der Schnittstelle sind Vielfältig. So können beispielsweise Smarte Ladesäulen den aktuellen Strommix kontinuierlich erfragen und ihren Ladevorgang den Umweltverhältnissen anpassen. Ladesäulen können so immer auf „grünen“, CO_2-armen Strom setzen. Auch ist es dank der Vorhersage denkbar, die Ladesäulen so zu schalten, dass sie einen Ladevorgang erst dann starten, wenn ein

besserer Strommix vorliegt. Zudem kann die Schnittstelle zur Visualisierung von Daten genutzt werden.

5.1 Webseite

Im Zuge der Softwareentwicklung ist eine Webseite entstanden, über welche sich die Daten des CO_2-Kompass verwerten und darstellen lassen[4]. Am Anfang der Webseite stellt ein sich stetig aktualisierendes Liniendiagramm inklusive gleitendem Mittelwert die Emission der deutschlandweiten Energieerzeugung für die letzten 12 Monate dar (siehe Abbildung 2). Durch eine Verschiebung des Cursors, der sich unterhalb des Liniendiagrammes befindet, kann der Zeitraum der betrachtet werden soll eingestellt werden. Dies ermöglicht es, eine genauere Auskunft über die Emissionen der letzten drei Tage zu bekommen. Fährt man den Mauszeiger nun über das Diagramm, wird die Emission für einen bestimmten Zeitpunkt (im 15-Minuten-Rhythmus) angezeigt.

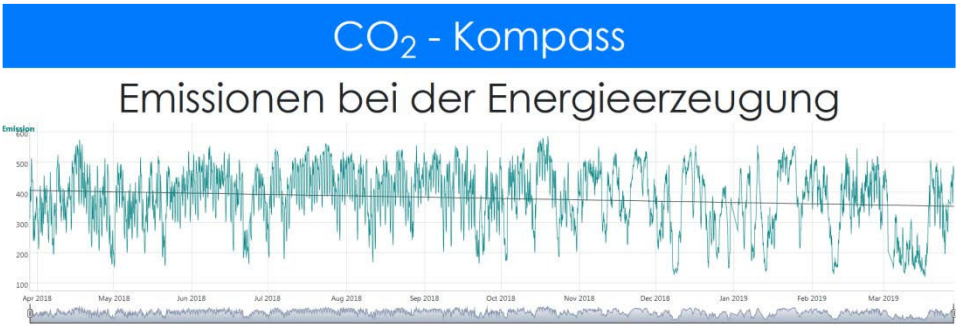


Abb. 2 - Emissionen bei der Energieerzeugung (Screenshot der Webseite, 12-Monats-Ansicht)

Der Webseite können neben den Emissionswerten auch Informationen über die aktuelle Stromproduktion pro Stromart in Megawatt und über den aktuellen prozentualen Anteil der regenerativen Energiequellen entnommen werden (siehe Abbildung 3). Diese Werte können je nach Einstellung für Deutschland im Gesamten oder für einen der vier Netzanbieter Transnet BW, TenneT GER, Amprion oder 50 Hertz angezeigt werden.

[4] www.co2compass.ml

Abb. 3 - Aktuelle Werte der Stromproduktion (Screenshots der Webseite)

5.2 Integration in Smarte Ladesäulen

Neben der Webseite, über die eine nutzerfreundliche Visualisierung der Daten und Berechnungen möglich ist, bietet der vorgestellte Prototyp weitere Potenziale für einen nutzenstiftenden Einsatz in praktischen Anwendungsfällen. Durch die Bereitstellung der aggregierten Daten und der entsprechenden Analyseergebnisse über eine REST-Schnittstelle können beliebige andere Dienste Zugriff erhalten und darauf aufbauend ihr Leistungsangebot verbessern, beispielsweise um Kunden einen umweltfreundlicheren Konsum von Elektrizität zu ermöglichen. Intelligente Lademanagementsysteme können so sicherstellen, dass Fahrzeugbatterien von Elektroautos zu Zeitpunkten geladen werden, an denen der Strommix möglichst klimaneutral oder kostengünstig [En18; Ke16; Ya15] ist. Durch die Vorhersagefunktionalität ist zusätzlich noch eine Abschätzung möglich, wann das Fahrzeug vollgeladen ist. So kann mit Hilfe einer App eingestellt werden, wie lange ein Fahrzeug geparkt wird und nach welchen Kriterien die Ladung (Emissionen, Kosten, Ladedauer, etc.) optimiert werden soll. Dies ermöglicht neben den zuvor genannten Potenzialen beispielsweise auch die Vermeidung von Lastverschiebungen für die Stromzufuhr [Pa12; Fl13]. Durch den Einsatz bidirektionaler Ladesäulen kann dieser Effekt darüber hinaus noch verstärkt werden. Sie ermöglichen eine Entladung der Autobatterie und damit eine Rückspeisung der Energie in das eigene Netz [Pe11], wodurch die Fahrzeuge als mobile Stromspeicher genutzt werden können. Im Sinne einer klimaneutralen Ausrichtung kann somit CO_2-armer Strom gespeichert und zu Zeitpunkten mit höherem CO_2-relevantem Primärenergieträger-Anteil genutzt werden. Damit trägt der Prototyp in mehrfacher Hinsicht positiv zu zukünftigen Entwicklungen bei der Digitalisierung des Energiesystems bei: So können zum einen Geschäftsmodelle näher an den Erwartungen und Bedürfnissen der Kunden ausgerichtet werden, zum anderen kann die ökologische Nachhaltigkeit weiter verbessert werden.

6 Zusammenfassung und Ausblick

Der CO_2-Kompass wurde vor dem Hintergrund entwickelt, Emissionen im Zuge der Stromproduktion transparenter darzustellen, so dass Endkunden eine nachhaltigere Energiezufuhr für ihre Produkte wählen können. Dabei bietet sich nicht nur die hier beschriebene Ladesäule als Anwendungsbeispiel an. Vielmehr kann Software über die REST-Schnittstelle mit jeder beliebigen Hardware verknüpft werden. Für zukünftige Ausarbeitungen würden sich hier zum Beispiel Maschinen die ein gewisses zeitliches Verschiebungspotenzial haben (z.B. Wärmepumpen, Kühllager etc.) anbieten. In Bezug auf die Verknüpfung mit Ladesäulen, stellt die Software inklusive Schnittstelle ein großes Entwicklungspotential dar. Durch die Entwicklung einer angepassten Benutzeroberfläche wäre es zum Beispiel möglich, eine Steuerung des Ladevorgangs über die vorhergesagten Emissionswerte zu realisieren. Um diese Entwicklungspotentiale voll auszuschöpfen, ist es wichtig für zukünftige Forschungsarbeiten eine Ladesäule als Prototyp mit der hergestellten Software auszustatten um Funktionalitäten zu testen und zu verbessern. Eine solche Integration von Dienstleistung und Produkt wurde in vielen wissenschaftlichen Ausarbeitungen als Product-Service System definiert und kann für einen erweiterten Fokus auf nachhaltige Lösungen sorgen [MO02; Hü18]. Daher bietet es sich für die zukünftige Forschung an, den CO_2-Kompass in ein Product-Service System einzubetten und mögliche Geschäftsmodelle für ein solches System zu erstellen.

Zudem sollten Limitationen dieser Arbeit bei zukünftigen Forschungsprojekten bedacht werden. Unter anderem wurde im Zuge des Scrum-Ansatzes ein Abgrenzungskriterium aufgestellt, welches als limitierender Aspekt dieser Arbeit aufgefasst werden kann. Um der Entwicklung des Tools einen ersten Rahmen zu geben wurde folgendes Abgrenzugskriterium aufgestellt: ‚Die Anpassung der Benutzeroberfläche der Ladesäulen, sowie das Steuern der Ladefunktion sind kein Bestandteil der Toolentwicklung'. Zukünftige Ausarbeitungen sollten das Ziel haben, den CO_2-Kompass in Ladesäulen zu integrieren. So können neue Erkenntnisse über das Tool erlangt werden und eine stetige Optimierung kann erfolgen.

7 Danksagung

Dieser Beitrag entstand im Rahmen des Forschungsprojektes „SmartHybrid – Electrical Engineering" (ID: ZW 6-85003732) das durch den Europäischen Fonds für regionale Entwicklung (EFRE) und das Land Niedersachsen (Investitions- und Förderbank Niedersachsen – NBank) finanziert wird. Wir bedanken uns bei den Förderern für die Unterstützung. Zudem gilt unser Dank Marvin Büchel, Lukas Mönck, Alexander Sprengel, Roman Schnell und Leon Frankenberg für deren Arbeit an der Entwicklung des CO_2-Kompass.

Literaturverzeichnis

[Br17] Brasseur, G. P., Jacob, D., & Schuck-Zöller, S. (2017). Klimawandel in Deutschland: Entwicklung, Folgen, Risiken und Perspektiven. Springer.

[En18] Ensslen, A., Ringler, P., Dörr, L., Jochem, P., Zimmermann, F., & Fichtner, W. (2018). Incentivizing smart charging: Modeling charging tariffs for electric vehicles in German and French electricity markets. *Energy research & social science*, *42*, 112-126.

[Fl13] Flath, C. M., Ilg, J. P., Gottwalt, S., Schmeck, H., & Weinhardt, C. (2013). Improving electric vehicle charging coordination through area pricing. Transportation Science, 48(4), 619-634.

[GL10] Gloger, B. (2010). Scrum. Informatik-Spektrum, 33(2), 195-200.

[Ha11] Hacker, F., Harthan, R., Kasten, P., Loreck, C., & Zimmer, W. (2011). Marktpotenziale und CO2-Bilanz von Elektromobilität.

[Hü18] Hüer, L., Hagen, S., Thomas, O., & Pfisterer, H. J. (2018). Impacts of product-service systems on sustainability–a structured literature review. Procedia CIRP, 73, 228-234.

[Ke16] Kempker, P., Dijk, N. V., Scheinhardt, W., Berg, H. V. D., & Hurink, J. (2016, January). Optimization of charging strategies for electric vehicles in PowerMatcher-driven smart energy grids. In Proceedings of the 9th EAI International Conference on Performance Evaluation Methodologies and Tools (pp. 242-249). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).

[MO02] Mont, O. K. (2002). Clarifying the concept of product–service system. Journal of cleaner production, 10(3), 237-245.

[NA14] Nationale Plattform Elektromobilität (2014) Fortschrittsbericht 2014 –Bilanz der Marktvorbereitung.http://nationale-plattform-elektromobilitaet.de/fileadmin/user_upload/Redaktion/NPE_Fortschrittsbericht_2014_Barrierefrei.pdf

[Or16] Orsi, F., Muratori, M., Rocco, M., Colombo, E., & Rizzoni, G. (2016). A multi-dimensional well-to-wheels analysis of passenger vehicles in different regions: Primary energy consumption, CO2 emissions, and economic cost. Applied Energy, 169, 197-209.

[Pa12] Paetz, A. G., Jochem, P., & Fichtner, W. (2012). Demand Side Management mit Elektrofahrzeugen–Ausgestaltungsmöglichkeiten und Kundenakzeptanz. In Symposium Energieinnovation (Vol. 15, No. 17.02, p. 2012).

[Pe11] Pehnt, M., Helms, H., Lambrecht, U., Dallinger, D., Wietschel, M., Heinrichs, H., ... & Behrens, P. (2011). Elektroautos in einer von erneuerbaren Energien geprägten Energiewirtschaft. Zeitschrift für Energiewirtschaft, 35(3), 221-234.

[Pl18] Plötz, P., Gnann, T., Wietschel, M., Kluschke, P., Doll, C., Hacker, F., ... & Lambrecht, U. (2018). Alternative Antriebe und Kraftstoffe im Straßengüterverkehr–Handlungsempfehlungen für Deutschland.

[Pr15] Prys-Hansen, M., Lellmann, M., & Röseler, M. (2015). Die Bedeutung der Klimafinanzierung für den Pariser Klimagipfel 2015

[Sc19] Schiffer, H. W. (2019). Zur energiewirtschaftlichen Notwendigkeit der Braunkohle für die Energieversorgung in Deutschland. *Zeitschrift für Energiewirtschaft*, 1-14.

[Sc15] Schill, W. P., Gerbaulet, C., & Kasten, P. (2015). Elektromobilität in Deutschland: CO2-Bilanz hängt vom Ladestrom ab. *DIW-Wochenbericht*, *82*(10), 207-215.

[Sc14] Schlömer, S., Bruckner, T., Fulton, L., Hertwich, E., McKinnon, A., Perczyk, D., ... & Wiser, R. (2014). Annex III: Technology-specific cost and performance parameters. Climate change, 1329-1356.

[SE13] Schlomann, B., & Eichhammer, W. (2013). Energieverbrauch und CO2-Emissionen industrieller Prozesstechnologien: Einsparpotenziale, Hemmnisse und Instrumente. T. Fleiter (Ed.). Fraunhofer-Verlag.

[SC97] Schwaber, K. (1997). Scrum development process. In Business object design and implementation (pp. 117-134). Springer, London.

[Th14] Thiel, C., Schmidt, J., Van Zyl, A., & Schmid, E. (2014). Cost and well-to-wheel implications of the vehicle fleet CO2 emission regulation in the European Union. Transportation Research Part A: policy and practice, 63, 25-42.

[Tr13] Trommer, S., Schulz, A., Hardinghaus, M., Gruber, J., Kihm, A., & Drogosch, K. (2013). Verbundprojekt Flottenversuch Elektromobilität–Teilprojekt Nutzungspotenzial. *Schlussbericht. Studie im Auftrag des Bundesministeriums für Umwelt, Naturschutz und Reaktorsicherheit (BMU). Berlin: Deutsches Zentrum für Luft-und Raumfahrt e. V.(DLR)*.

[Um17] Umwelt Bundesamt, UBA (2017). Nettostromerzeugung in Deutschland 2016 nach Primärenergieträgern (entnommen aus: https://www.umweltbundesamt.de/sites/default/files/medien/372/bilder/dateien/stromm ix-karte-2016.pdf)

[Um19] Umwelt Bundesamt, UBA (2019). Erneuerbare Energien - Vermiedene Treibhausgase, 15.03.2019 (entnommen aus: https://www.umweltbundesamt.de/daten/energie/erneuerbare-energien-vermiedene-treibhausgase)

[Ya15] Yang, H., Yang, S., Xu, Y., Cao, E., Lai, M., & Dong, Z. (2015). Electric vehicle route optimization considering time-of-use electricity price by learnable partheno-genetic algorithm. IEEE Transactions on Smart Grid, 6(2), 657-666.

[Ze19] Zellner, R. (2019). Zu viel CO2 aus dem Verkehr: Ist Elektromobilität die Lösung?. Nachrichten aus der Chemie, 67(3), 26-31.

[Zi16] Zimmer, W., Blanck, R., Bergmann, T., Mottschall, M., Waldenfels, R.; Förster, H. et al. (2016): Endbericht Renewbility III. Optionen einer Dekarbonisie-rung des Verkehrssektors. Studie im Auftrag des BMUB 2016. Öko-Institut; DLR; ifeu Insti-tut für Energie-und Umweltforschung Heidelberg (IFEU); Infras

Overview of machine learning and data-driven methods in agent-based modeling of energy markets

Ashreeta Prasanna [1], Sascha Holzhauer[2] and Friedrich Krebs[3]

Abstract: Local energy markets (LEM) allow prosumers and consumers to trade energy directly between one another and offer flexibility services to the grid. The benefits and challenges of such LEM need to be identified, and agent-based modeling (ABM) is a useful method to conduct simulation experiments that compare different market structures and clearing mechanisms. Machine learning (ML) and data-driven methods when integrated with ABM show great potential for constructing new distributed, agent-level knowledge. In this paper, we discuss the requirements for coupling ML methods and ABM. We also provide an overview of published literature on the common methods of integration of ML and data-driven methods in ABM and discuss how these requirements are commonly addressed.

Keywords: machine learning, agent-based modeling, local energy markets, reinforcement learning, load forecasting

1 Introduction

The widespread adoption of renewable energy supply technologies and the availability of data from smart meters and other monitoring systems allows the development of local energy markets (LEM), also referred to as peer-to-peer markets or direct energy markets. LEM aim to offer multiple benefits: implementation of prosumers' preferences: for instance for renewable energy or lower CO_2 emissions; reduction in energy costs; reduction in costs for grid investment; and flexible and efficient locally managed energy supply [Fa14, So18]. Price signals that indicate scarcity and excess of fluctuating energy supply could incentivize prosumers to act beneficially for the energy system. A number of projects use demonstration and/or modeling to analyze the benefits and drawbacks of LEM and their design [BOR17, Mo18, Me18, RKF16, RM13, So18, Zh18]. ABM is found to be particularly suitable to evaluate the design of LEMs because it allows the representation of aspects such as learning effects in repeated interactions, asymmetric information, imperfect competition, or strategic interaction and collusion in a more realistic way [RKF16, Se07].

[1] University of Kassel, Department Integrated Energy Systems, Wilhelmshöher Allee 73, 34121 Kassel, Germany, ashreeta.prasanna@uni-kassel.de
[2] University of Kassel, Department Integrated Energy Systems, Wilhelmshöher Allee 73, 34121 Kassel, Germany, sascha.holzhauer@uni-kassel.de
[3] University of Kassel, Department Integrated Energy Systems, Wilhelmshöher Allee 73, 34121 Kassel, Germany, friedrich.krebs@uni-kassel.de

doi:10.18420/inf2019_73

As agent-based models (ABM) are data intensive, automating or semi-automating the process of capturing system knowledge using ML and other data-driven methods is a growing field of research. This is especially true for LEM when agents interact with the dynamic energy system and time constraints need to be considered in forecasting market prices, energy consumption and generation as well in the bidding process.

ML algorithms can be classified into three broad categories: supervised learning, unsupervised learning and reinforcement learning [Al10]. Supervised learning algorithms are used to develop a predictive model based on both input and output data. Some examples of supervised learning algorithms are k-Nearest neighbors, support vector machines, decision trees, neural networks, etc. Unsupervised learning algorithms are used to group and interpret data based only on input data. Common unsupervised learning algorithms are k-means clustering, hierarchical clustering, DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering, etc. Reinforcement learning (RL) is a goal directed approach where an agent learns the optimal behavior through repeated trial-and-error interactions with the environment without human involvement. Examples of RL algorithms are Q-learning, genetic algorithms, Erev-Roth reinforcement learning, learning classifier systems, etc.

ML can be coupled with ABM in a number of different ways (see figure 1). One possibility is to use ML in forecasting external input data for the ABM, which can then subsequently be used to inform agent behavior [FPV16, Pi16a, Pi16b, Sa16]. This is shown in the top half of figure 1, where ML is used to forecast aspects such as production, load and market price and provide these inputs to the agent. The second possibility is to use ML algorithms to implement the learning behavior of agents when they place bids on the market [KUP03, MGW18, Pe13]. This is shown in the bottom half of figure 1 where the learning behavior of agents could be either rule-based (the predefined strategies in figure 1) or through using RL algorithms.

It is also possible to use supervised learning techniques (as an alternative to RL) in a two-step approach to allow agents to place bids. Fischer [Fi18] describes this approach for financial markets where supervised learning is first used to build a predictive model using historical data, and then the forecasts from this predictive model are fed into a trading module to derive the trading action, e.g. buy or sell when the forecasted market price passes a certain threshold. There are a number of limitations in using supervised learning to directly place bids, which are discussed by Fisher [Fi18]. First, the optimization objective in the predictive model, i.e., the minimization of the forecast error, is not necessarily in line with the ultimate goal of the agent, e.g., the maximization of profits. Second, in most cases, only the forecast itself is used as an input, and additional valuable information that could be obtained from the feature space is discarded [Fi18, Mo98]. Finally, in the context of ABM with a large number of agents that interact dynamically, it is desirable to use lean algorithms that are computationally efficient. The use of RL as an alternative to supervised learning allows the forecast and the subsequent selection of a strategy to be carried out in one single step and both to be optimized in line with the objective of the agent [Fi18]. Therefore, RL and novel implementations of RL such as

multi-agent RL and deep reinforcement learning (deep RL) are popular solutions to implement learning behavior in agents.

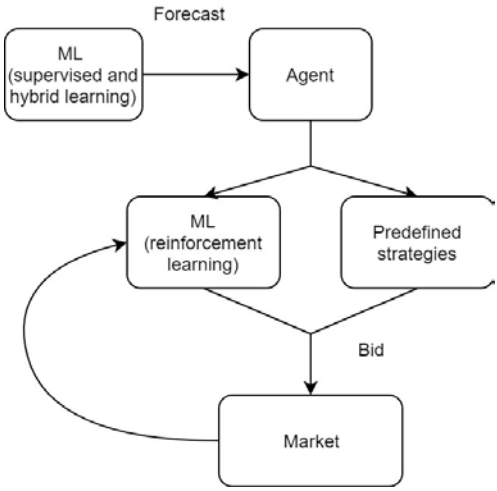


Figure 1: Possible use of ML and data-driven methods in ABM of energy markets

This paper aims to provide an overview of published literature on the common methods of integration of ML and data-driven methods in ABM of energy markets. While there are a number of published reviews on the use of ML for forecasting, e.g. [HF16] or on RL methods to improve decision making in agents, e.g. [WV08], there is no comprehensive overview on how different ML and data-driven methods can help improve ABM of LEM and the specific requirements for their integration.. In this paper, published journal articles (from year 2000 onwards) on integration of RL methods in ABM of energy markets are selected and presented in Section 1. In Section 2, we discuss the requirements of developing ML and data-driven methods for integration with ABM making reference to this selected literature. The conclusions are presented in the final section.

2 Brief summary of published literature on machine learning used in agent-based models of energy markets

Recent literature on ABM of energy markets includes the use of ML and other data based approaches to improve models and represent complexity in simulations. Selected journal articles on ABM of energy markets that include ML and data based methods are presented in table 1. For each reference, if ML is used in forecasting certain values, the subject of forecast, for example electricity market price, renewable generation, load, etc., is identified and noted in the second column of table 1. The type of ML algorithm used to derive or calculate the subject of forecast is noted in the third column of table 1. If RL is used, the kind of learning algorithm is identified and noted in the last column. The

objective of RL algorithms in the selected references is to place bids on the market. Most references describe either the use of ML for forecasting, or RL for agents' bids on the market, however some references mention both cases. Even if forecasts are used as input data to the learning algorithm or to model agent behavior, the details of the forecasted data and the methods used for forecasting might not be specified in the article. In this case, the comment 'not specified' is noted in the relevant column.

2.1 Forecasting to inform decision-making

The bidding behavior of agents in the context of electricity or energy markets is often informed by the forecasted values of a number of inputs such as the forecasted load, generation and market price. Short-term forecasting methods are the most relevant when considering wholesale day-ahead or intra-day energy markets. Supervised learning methods and statistical methods are the most commonly used ML methods in forecasting [HF16]. In addition to forecasting based on historical weather, load and market price, the introduction of smart meters in many markets also provides a valuable source of more detailed data for forecasting loads [De11, Wa18]. In some cases unsupervised methods such as clustering are applied along with supervised learning algorithms or statistical methods to provide forecasts [Au18, FPV16, MW18].

A number of articles have reviewed ML methods for forecasting; however, these reviews do not consider the use of these forecasts for ABM or simulations. Table 1 includes some published articles where forecasting methods have been specifically developed for integration in ABM of energy markets. However, only few articles elaborate on the methods they use to derive the parameters that are used to inform the bidding behavior of agents. Since this is a new area of research, there is scope for further research on selection of ML algorithms for the particular case for forecasting as an input to ABM.

2.2 Multi-agent reinforcement learning for intelligent bidding

As discussed in the previous section, RL is a popular solution to implement learning behavior in agents. Although the agents can be endowed with behaviors designed in advance, they often need to learn new behaviors online such that the performance of the agent or of the whole multi-agent system gradually improves [Bu10, SW99, SV00]. In the case of electricity or energy markets, since the environment changes over time, a hardwired or pre-defined behavior of agents is inappropriate. The articles in table 1 apply a variety of RL algorithms with single or multiple agents to conduct experiments of different types of electricity market simulations. A broad spectrum of RL algorithms exist, e.g., model-free methods based on the online estimation of value functions, model-based methods (typically called dynamic programming), and model-learning methods that estimate a model, and then learn using model-based techniques [Bo10]. In the selected literature, mainly model free approaches have been applied, for example Q-

learning and SARSA e.g. [Bo18, BEC18, EKS17, KUP03, PRD18, Pe13, Ya18] but some authors have also used model-based techniques, e.g. [BO01, VI08, Zh16].

In addition to the learning approach, another consideration is the definition of an appropriate formal goal for the learning. The articles in table 1 mainly focus on cases where agents act non-cooperatively to maximize their own interests. Mguni et al. find that while the lack of coordination produces stable outcomes or Nash equilibria, these are vastly suboptimal from a system perspective [Du08, Mg19]. Therefore, they propose an incentive-design method that modifies agents' rewards in a non-cooperative ABM that results in independent, self-interested agents choosing actions that produce optimal system outcomes in strategic settings.

Experience sharing, for instance agents exchanging information using communication, skilled agents serving as teachers for the learner, or the learner watching and imitating the skilled, can help agents with similar tasks learn faster and reach better performance [Bu10]. However, in the selected studies, there is no direct information exchange between agents, and the information flow is mainly directed from the market to the agents. Most of the studies consider the case of agents competing to maximize their own profits under different levels of market competition, however, Zhang et al. [Zh17] consider the case of optimal consensus control. Therefore, in the context of providing flexibility and encouraging consumption of electricity produced locally, it might be relevant to consider cases where the agents (prosumers and consumers) pursue cooperative strategies rather than purely competing strategies.

Reference	**Forecast: subject**	**Forecasting algorithm/ Derivation method**	**Learning algorithm**
Faia et al., 2016 [FPV16]	Electricity price in contracts	Hybrid (k-means and fuzzy logic)	Not specified
Aliabadi et al., 2017 [EKS17]	Locational marginal price at each node	DC- Optimal power flow problem	Q-learning
Pinto et al., 2016 [Pi16b]	Electricity market price	Support Vector Machines	Not specified
Kutschinski et al., 2003 [KUP03]	-	Not specified	Q-learning
Azadeh et al., 2010 [ASM10]	-	Not specified	Ant colony optimization
Zhang et al., 2017	-	Not specified	Adaptive dynamic

[Zh17]			programming
Mengelkamp at al., 2018 [MGW18]	Load	Standard profiles with error function	Erev-Roth reinforcement learning
Bunn & Oliveira, 2001 [BO01]	-	Not specified	Defined strategies
Visudhiphan & Ilic, 2008 [VI08]	-	Not specified	Defined strategies
Zhou et al., 2011[ZZW11]	Load, electricity price	Simulation model, polynomial cost function (producer)	Erev-Roth reinforcement learning
Yu et al., 2019 [Yu19]	-	Not specified	Experience-weighted attraction learning
Viehmann et al, 2018 [VLM18]	-	Not specified	Q-learning
Peters et al., 2013 [Pe13]	-	Not specified	State-Action-Reward-State-Action (SARSA)
Yang et al., 2018 [Ya18]	Load	k-means clustering	Q-learning
Patyn et al., 2018 [PRD18]	-	Not specified	Fitted Q-iteration with: a multilayer perceptron, a convolutional neural network and a long short-term memory neural network
Boukas et al., 2018 [BEC18]	-	Not specified	Q-learning, Q-function with a Neural Network
Boukas et al., 2018 [Bo18]	-	Not specified	Q-learning, Deep Q-Network
Chen et al., 2019 [CLS19]	Electricity market price	Extreme Machine Learning	Not specified

Tab. 1: A selection of published articles which use ML and data based methods in ABM of energy markets.

3 Requirements for the integration of machine learning and data-driven methods in multi-agent systems

3.1 Computational efficiency

Low computational demands mean lower costs, which increases the likelihood of automated bidding agents based on ML algorithms being deployed at prosumer's premises. In the selected literature, performance comparisons which include computational efficiency between different variations on algorithms which use the same RL approach are presented. For example, Patyn et al. [PRD18] use model-free RL to model the a heat pump agent which shifts loads in a day-ahead market to minimize daily electricity costs. They approximate the Q-function by three different neural architectures, a multilayer perceptron (MLP), a convolutional neural network (CNN) and a long short-term memory neural network (LSTM), and find that all architectures outperform a trivial thermostat controller and shift loads successfully after 20-25 days. In their modeled case, they do not find a significant difference in the performance of the MLP and the LSTM, both of which outperform the CNN model. However, they find that the MLP requires far less computation time. Pinto et al. [Pi16b] compare a support vector machines (SVM) based approach with artificial neural networks (ANN) to forecast the electricity market price. They show SVM methods provide similar results but take half the time of ANN. Finally, Mengelkamp et al. [MGW18] find the computational time for RL based strategies to be twice as high compared to bidding with random prices or with a selected fixed price. However, the computational time of their implementation of different variations of RL strategies differ by only 6%. Thus, they do not consider computational time as a criterion for selecting a particular strategy.

Deep RL or the use of deep neural networks within RL for value function approximations has also been shown to be successful in is in scaling up prior work in RL to high-dimensional problems. By means of representation learning, they can deal efficiently with the curse of dimensionality, unlike tabular and traditional non-parametric methods [Ar17, BCV13]. A relevant future research direction would be to compare the performance of dynamic programming approaches with deep RL approaches, since these are state of the art RL algorithms. The availability of open source implementations of different reinforcement algorithms (discussed in section 3.2) allows for the definition of standard benchmarks for testing new algorithms and evaluating new techniques in a standardized manner.

3.2 Learning curve or difficulty of implementation of machine learning methods

The learning curve in implementing ML methods is an important consideration because ABM developers cannot focus exclusively on the implementation of these methods but also need to consider other aspects of modeling such as interactions between agents and the mechanics of market clearing. Therefore, the availability of standard libraries, examples and detailed documentation are a consideration when selecting the method for implementation. While most publications do not mention the details and the use of standard libraries used in their implementation of ML algorithms, a wide selection of open source libraries are available in common programming languages to implement supervised, unsupervised, and RL algorithms.

Some common libraries in Python to implement RL are OpenAI Gym or Universe, RLLib, Coach, TensorForce, Keras-RL, PyBrain, RLPy [Ge13, In19, Op19, Pl16, Ra19, Re19, Sc10]. Libraries implemented in Java for RL are BURLAP, RL4J, RL-Glue [Ch19a, Sk19, TW09] and packages for R are ReinforcementLearning and MDPtoolbox [CH19b, PF19].

MATLAB also offers a number of libraries to implement ML algorithms, Pinto et al. [Pi16b] use it to develop their SVM approach to forecast market prices and Mengelkamp et al. [MGW18] use it to implement RL algorithms. Chen & Su [CS18] implement their RL algorithm in Python, and Lamperti et al. [La18] also use Python to implement their model calibration approach.

In addition to standardized frameworks for implementation, another consideration with respect to the difficulty of implementation is the definition of an appropriate formal goal for the learning multi-agent system. As discussed in Section 2.2, a common approach is to apply single-agent Q-learning to the multi-agent case where the learned Q-functions only depend on the current agent's action without being aware of the other agents. Busoniu et al. [Bu10], find that one important research direction is understanding the conditions under which single-agent RL works in mixed stochastic games, especially given the preference towards using single-agent techniques for multi-agent systems in practice.

3.3 Flexibility or adaptability of the machine learning algorithms in a multi-agent system

ML models can have different learning rates with different datasets, and need to be tuned so that they can optimally solve the ML problem. The measures used to tune a model are called hyperparameters. In the context of ML providing input data to an agent in an ABM, it is important that the hyperparameters can be easily set and adjusted to allow selection between accuracy and computational time, for example. In the Multi-Agent System for Competitive Electricity Markets (MASCEM) platform developed by Santos et al., the management of the system to adapt its execution time to the purpose of the simulation is performed by means of a fuzzy process [Sa16]. Standard libraries in Py-

thon, for example scikit-learn, allow hyperparameter optimization using several methods like grid search, random search and Bayesian optimization. These methods could be integrated in the architecture of the ABM platform to enable adaptability of the ML algorithms.

In the context of RL algorithms, the algorithms can be tuned by choosing the learning rate, selecting the resolution of the value function, choosing how often to update the representation of the value function, and making tradeoffs between exploring to improve the learning model and exploring to improve the learning policy [AS02]. The consequences of these choices are greatly influenced by which RL approach is selected and the specific details of how the algorithm is implemented. Atkeson and Santamaria [AS02] find that there are fewer parameter choices to make in model-based RL. The (hyper) parameter values in RL also influence whether convergence is achieved and how quickly it is achieved.

3.4 Robustness

The agent's perception of the environment may vary, and therefore the robustness of an ML algorithm is an important consideration. Multi-agent RL is inherently robust because if one or more agents fail in a multi-agent system, the remaining agents can take over some of their tasks [Bu10]. Other properties of multi-agent RL are stability and adaptation: an opponent-independent algorithm converges to a strategy that is part of an equilibrium solution regardless of what the other agents are doing while an opponent-aware algorithm learns models of the other agents and reacts to them using some form of best response. Algorithms focused on stability (convergence) only are typically unaware and independent of the other learning agents [Bu10]. Common methods to measure robustness are convergence time and change in output/convergence values across multiple runs. In Rosen & Madlener [RM13], tests which consider the speed of convergence are used to quantify robustness of the algorithm. In Viehmann et al. [VLM18] each model is run multiple times with varying seeds to check for multiple stable outcomes and robustness of results. Peters et al. [Pe13] consider noise injection, to alleviate overfitting and improve generalization in supervised settings.

None of the selected articles compares the robustness of different algorithms. However, a general understanding is that ABM with agents that are unaware or do not directly interact with the other agents converge more easily, while in other cases reward functions or other criteria need to be specifically defined in order to achieve convergence. For example, Zhou et al. [ZWL18] use step length control and learning process involvement to facilitate convergence and also define a last-defense mechanism (ending the simulation after a pre-defined finite number of iterations, regardless if convergence is achieved or not) to handle divergence.

4 Conclusions

In this paper, we provide an overview of published literature on the common methods of integration of ML and data-driven methods in ABM of energy markets. We discuss some important requirements for this integration and present the methods used in published articles to address these requirements.

Since the integration of ML methods in ABM is a relatively new area of research, there are few articles which discuss the methods of such integration and the benefits it can offer. The purpose of our contribution is to provide a first (to the best of our knowledge) review of such novel approaches which may serve as a starting point for future research efforts.

Further case studies are required for a clear comparison considering the highlighted dimensions as well as additional dimensions. As discussed in section 2.1, further research on the selection of suitable and efficient ML algorithms specifically to provide inputs for ABM and simulations is necessary. In addition, a formalized architecture and a common module which can use data inputs, e.g. weather related and load related parameters such as temperature-humidity index, wind chill index, etc. used for the forecasting algorithms would improve the efficiency and modularity of integrating ML with the ABM.

With respect to implementing learning behavior in agents, a number of future research areas have been identified: comparison of the robustness of different algorithms, the suitability and selection of RL algorithms specific to the use case of bidding on markets, and identification of algorithms which can better represent cooperative strategies, and conversely, non-cooperative agent strategies. Finally, it is also relevant to conduct experiments where the agents (prosumers and consumers) pursue cooperative strategies rather than purely competing strategies. In the reviewed literature, it is difficult to compare the efficiency of different RL algorithms because they have been implemented in different types of ABM, with different assumptions and market dynamics. Experiments that compare different RL approaches but with the same market assumptions would be valuable, as they would help in benchmarking the different algorithms and provide the possibility to identify which algorithms are more suited for or efficient in specific market designs.

In summary, ML can be used in ABM, for example to forecast input parameters which agents can use in their decision making, and, for the learning as the simulation goes along (i.e. for RL). There is scope for further research and definition of standardized test cases on all types of coupling, as well as definition of standardized methods to evaluate new techniques. Numerous open source libraries and frameworks allow such implementation to be feasible and efficient.

Bibliography

[Al10] Alpaydin, E.: Introduction to Machine Learning. 2nd. ed.: The MIT Press, 2010.

[Ar17] Arulkumaran, K. et al.: Deep reinforcement learning: A brief survey. In: IEEE Signal Processing Magazine vol. 34, Nr. 6, pp. 26–38, 2017.

[AS02] Atkeson, C.G. ; Santamaria, J.C.: A comparison of direct and model-based reinforcement learning, Nr. April, pp. 3557–3564, 2002.

[Au18] Auder, B. et al.: Scalable Clustering of Individual Electrical Curves for Profiling and Bottom-Up Forecasting. In: Energies vol. 11, Nr. 7, p. 1893, 2018.

[ASM10] Azadeh, A.; Skandari, M.; Maleki-Shoja, B.: An integrated ant colony optimization approach to compare strategies of clearing market in electricity markets: Agent-based simulation. In: Energy Policy vol. 38, Elsevier, Nr. 10, pp. 6307–6319, 2010.

[BCV13] Bengio, Y.; Courville, A.; Vincent, P.: Representation learning: A review and new perspectives. In: IEEE Transactions on Pattern Analysis and Machine Intelligence vol. 35, Nr. 8, pp. 1798–1828, 2013.

[Bo18] Boukas, I. et al.: Intra-day Bidding Strategies for Storage Devices Using Deep Reinforcement Learning. In: International Conference on the European Energy MarketISBN 9781538614884, 2018.

[BEC18] Boukas, I.; Ernst, D.; Cornelusse, B.: Real-Time Bidding Strategies from Micro-Grids Using Reinforcement Learning. In: CIRED Workshop 2018, Nr. 0440, pp. 7–8, 2018.

[BOR17] Bremdal, B.A.; Olivella, P.; Rajasekharan, J.: EMPOWER: A network market approach for local energy trade. In: 2017 IEEE Manchester PowerTech, 2017, pp. 1–6

[BO01] Bunn, D. ; Oliveira, F.S.: Agent-Based Simulation—An Application to the New Electricity Trading Arrangements of England and Wales. In: Ieee Transactions on Evolutionary Computation, vol. 5, Nr. 5, pp. 493–503, 2001.

[Bu10] Busoniu, L. et al.: Multi-agent reinforcement learning: An overview. In: (Srinivasan, D.; Jain, L. C. eds.): Innovations in Multi-Agent Systems and Applications -- 1, Studies in Computational Intelligence. vol. 310. Berlin, Germany : Springer, 2010, pp. 183–221

[Ch19a] Chades, I. et al.: Brown-UMBC Reinforcement Learning and Planning (BURLAP). http://burlap.cs.brown.edu/faq.html#cite. - accessed 22/4/2019

[Ch19b] Chades, I. et al.: MDPtoolbox: Markov Decision Processes Toolbox. https://cran.r-project.org/package=MDPtoolbox, accessed: 22/4/2019

[CLS19] Chen, K.; Lin, J.; Song, Y.: Trading strategy optimization for a prosumer in continuous double auction-based peer-to-peer market: A prediction-integration model. In: Applied Energy vol. 242, pp. 1121–1133, 2019.

[CS18] Chen, T. ; Su, W.: Indirect Customer-to-Customer Energy Trading with Reinforcement Learning. In: IEEE Transactions on Smart Grid vol. PP, IEEE, Nr. c, p. 1, 2018.

[De11] De Silva, D. et al.: Semi-supervised classification of characterized patterns for demand forecasting using smart electricity meters. In: 2011 International Conference on

Electrical Machines and Systems, ICEMS 2011, IEEE, pp. 1–6ISBN 9781457710445, 2011.

[Du08] Dubey, P.: Inefficiency of Nash Equilibria. In: Mathematics of Operations Research vol. 11, Nr. 1, pp. 1–8, 2008.

[EKS17] Esmaeili Aliabadi, D.; Kaya, M.; Sahin, G.: Competition, risk and learning in electricity markets: An agent-based simulation study. In: Applied Energy vol. 195, Elsevier Ltd, pp. 1000–1011, 2017.

[Fa14] Faber, I. et al.: Micro-energy markets: The role of a consumer preference pricing strategy on microgrid energy investment. In: Energy vol. 74, Elsevier Ltd, Nr. C, pp. 567–575, 2014.

[FPV16] Faia, R.; Pinto, Z.; Vale, Z.: Dynamic Fuzzy Clustering Method for Decision Support in Electricity Markets Negotiation. In: ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal vol. 5, Nr. 1, p. 23, 2016.

[Fi18] Fischer, T.G.: Reinforcement learning in financial markets-a survey, 2018

[Ge13] Geramifard, A. et al.: RLPy: The Reinforcement Learning Library for Education and Research, 2013.

[HF16] Hong, T. ; Fan, S.: Probabilistic electric load forecasting: A tutorial review. In: International Journal of Forecasting vol. 32, Elsevier B.V., Nr. 3, pp. 914–938, 2016.

[In19] Intel AI Lab: Reinforcement Learning Coach. https://nervanasystems.github.io/coach/., accessed: 22/4/2019

[KUP03] Kutschinski, E.; Uthmann, T.; Polani, D.: Learning competitive pricing strategies by multi-agent reinforcement learning. In: Journal of Economic Dynamics and Control vol. 27, Nr. 11–12, pp. 2207–2218, 2003.

[La18] Lamperti, F. et al.: Agent-based model calibration using machine learning surrogates. In: Journal of Economic Dynamics and Control vol. 90, pp. 366–389, 2018.

[Me18] Mengelkamp, E. et al.: Designing microgrid energy markets: A case study: The Brooklyn Microgrid. In: Applied Energy vol. 210, pp. 870–880, 2018.

[MGW18] Mengelkamp, E.; Gärttner, C.; Weinhardt, C.: Intelligent Agent Strategies for Residential Customers in Local Electricity Markets, pp. 97–107ISBN 9781450357678, 2018.

[MW18] Mengelkamp, E.; Weinhardt, C.: Clustering Household Preferences in Local Electricity Markets, pp. 538–543ISBN 9781450357678, 2018.

[Mg19] Mguni, D. et al.: Coordinating the Crowd: Inducing Desirable Equilibria in Non-Cooperative Systems, Nr. Id, 2019.

[Mo98] Moody, J. et al.: Performance functions and reinforcement learning for trading systems and portfolios. In: Journal of Forecasting vol. 17, Nr. December 1997, pp. 441–470, 1998.

[Mo18] Morstyn, T. et al.: Using peer-to-peer energy-trading platforms to incentivize prosumers to form federated power plants. In: Nature Energy vol. 3, Springer US, Nr. 2, pp. 94–101, 2018.

[Op19] OpenAI: *Gym*. https://gym.openai.com/., accessed: 22/4/2019

[PRD18] Patyn, C.; Ruelens, F.; Deconinck, G.: Comparing neural architectures for demand response through model-free reinforcement learning for heat pump control. In: 2018 IEEE International Energy Conference, ENERGYCON 2018, IEEE, pp. 1–6ISBN 9781538636695, 2018.

[Pe13] Peters, M. et al.: A reinforcement learning approach to autonomous decision-making in smart electricity markets. In: Machine Learning vol. 92, Nr. 1, pp. 5–39, 2013.

[Pi16a] Pinto, T. et al.: Adaptive Portfolio Optimization for Multiple Electricity Markets Participation. In: IEEE Transactions on Neural Networks and Learning Systems vol. 27, Nr. 8, pp. 1720–1733, 2016.

[Pi16b] Pinto, T. et al.: Support Vector Machines for decision support in electricity markets' strategic bidding. In: Neurocomputing vol. 172, Elsevier, pp. 438–445, 2016.

[Pl16] Plappert, M.: Keras-RL. https://keras-rl.readthedocs.io/en/latest/., accessed: 22/4/2019

[PF19] Pröllochs, N. ; Feuerriegel, S.: Reinforcement Learning in R. https://cran.r-project.org/web/packages/ReinforcementLearning/vignettes/ReinforcementLearning.html. - accessed 22/4/2019

[Ra19] The Ray Team Revision: RLlib: Scalable Reinforcement Learning. https://ray.readthedocs.io/en/latest/rllib.html., accessed: 22/4/2019

[Re19] reinforce.io: TensorForce - modular deep reinforcement learning in TensorFlow. https://tensorforce.readthedocs.io/en/latest/., accessed: 22/4/2019

[RKF16] Ringler, P. ; Keles, D. ; Fichtner, W.: Agent-based modelling and simulation of smart electricity grids and markets - A literature review. In: Renewable and Sustainable Energy Reviews vol. 57, Elsevier, pp. 205–215ISBN 1364-0321, 2016.

[RM13] Rosen, C. ; Madlener, R.: An auction design for local reserve energy markets. In: Decision Support Systems vol. 56, Elsevier B.V., Nr. 1, pp. 168–179, 2013.

[Sa16] Santos, G. et al.: MASCEM: Optimizing the performance of a multi-agent system. In: Energy vol. 111, Elsevier Ltd, pp. 513–524, 2016.

[Sc10] Schaul, T. et al.: PyBrain. In: Journal of Machine Learning Research, 2010.

[SW99] Sen, S. ; Weiss, G.: Learning in Multiagent Systems. In: Weiss, G. (ed.): Multiagent systems: A modern approach, Cambridge, MA, USA : MIT Press, 1999 — ISBN 0-262-23203-0, pp. 259–298

[Se07] Sensuß, F. et al.: Agent-based Simulation of Electricity Markets - A Literature Review. In: Energy Studies Review vol. 15, Nr. 2, p. 44, 2007.

[Sk19] Skymind: Deep Learning for Java. https://deeplearning4j.org/., accessed: 22/4/2019

[So18] Sousa, T. et al.: Peer-to-peer and community-based markets: A comprehensive review, 2018.

[SV00] Stone, P. ; Veloso, M.: Multiagent systems: a survey from a machine learning perspective. In: Autonomous Robots vol. 8, Nr. 3, pp. 345–383ISBN 0929-5593, 2000.

[TW09] Tanner, B. ; White, A.: RL-Glue: Language-Independent Software for Reinforcement-

Learning Experiments. In: J. Mach. Learn. Res. vol. 10, pp. 2133–2136, 2009.

[VLM18] Viehmann, J.; Lorenczik, S.; Malischek, R.: Multi-unit Multiple Bid Auctions in Balancing Markets : an Agent-based Q-learning Approach, Nr. 18, 2018.

[VI08] Visudhiphan, P. ; Ilic, M.D.: Dynamic games-based modeling of electricity markets, pp. 274–281 vol.1ISBN 0780344030, 2008.

[Wa18] Wang, Y. et al.: Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. In: IEEE Transactions on Smart Grid vol. 3053, Nr. June 2017, pp. 1–24, 2018.

[WV08] Weidlich, A. ; Veit, D.: A critical survey of agent-based wholesale electricity market models. In: Energy Economics vol. 30, Nr. 4, pp. 1728–1759ISBN 0140-9883, 2008.

[Ya18] Yang, Y. et al.: Recurrent Deep Multiagent Q-Learning for Autonomous Agents in Future Smart Grid Extended Abstract. In: Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), pp. 2136–2138, 2018.

[Yu19] Yu, Q. et al.: The Strategy Evolution in Double Auction Based on the Experience-Weighted Attraction Learning Model. In: IEEE Access vol. 7, pp. 16730–16738, 2019.

[Zh16] Zhang, C. et al.: A Bidding System for Peer-to-Peer Energy Trading in a Grid-connected Microgrid. In: Energy Procedia vol. 103, Elsevier B.V., Nr. April, pp. 147–152, 2016.

[Zh17] Zhang, H. et al.: Data-Driven Optimal Consensus Control for Discrete-Time Multi-Agent Systems with Unknown Dynamics Using Reinforcement Learning Method. In: IEEE Transactions on Industrial Electronics vol. 64, IEEE, Nr. 5, pp. 4091–4100, 2017.

[Zh18] Zhang, C. et al.: Peer-to-Peer energy trading in a Microgrid. In: Applied Energy vol. 220, Elsevier, Nr. December 2017, pp. 1–12, 2018.

[ZWL18] Zhou, Y.; Wu, J.; Long, C.: Evaluation of peer-to-peer energy sharing mechanisms based on a multiagent simulation framework. In: Applied Energy vol. 222, Elsevier, Nr. February, pp. 993–1022ISBN 0046405003, 2018.

[ZZW11] Zhou, Z.; Zhao, F.; Wang, J.: Agent-based electricity market simulation with demand response from commercial buildings. In: IEEE Transactions on Smart Grid vol. 2, IEEE, Nr. 4, pp. 580–588, 2011.

Influences in Forecast Errors for Wind and Photovoltaic Power: A Study on Machine Learning Models

Jens Schreiber, Artjom Buschin, Bernhard Sick[1]

Abstract: Despite the increasing importance of forecasts of renewable energy, current planning studies only address a general estimate of the forecast quality to be expected and selected forecast horizons. However, these estimates allow only a limited and highly uncertain use in the planning of electric power distribution. More reliable planning processes require considerably more information about future forecast quality. In this article, we present an in-depth analysis and comparison of influencing factors regarding uncertainty in wind and photovoltaic power forecasts, based on four different machine learning (ML) models. In our analysis, we found substantial differences in uncertainty depending on ML models, data coverage, and seasonal patterns that have to be considered in future planning studies.

Keywords: uncertainty analysis, machine learning models, seasonal effects, data coverage

1 Introduction

With the further expansion of wind and photovoltaic (PV) energy, the power supply system will change significantly in the coming decades. The overall power supply will become more weather-dependent and solutions must be found to ensure a robust and inexpensive power supply that maintains grid stability.

The major challenges for the grid stability caused by the energy system's transformation can mainly be traced back to two aspects: Firstly, the actual power supply of wind and solar energy plants is directly dependent on the weather and thus not directly compatible to consumption. Secondly, the expected power of the next hours and days is uncertain due to the strong dependence on the weather and must be predicted by power forecasts based on numerical weather predictions (NWPs).

Despite the increasing importance of forecasts for renewable power supply, current planning studies only address the forecast quality to be expected in the future for the whole of Germany based on representative forecasts (see, e.g., the dena grid study [KAS10]). Further, often, these studies only consider a limited number of forecasts horizons. However, these estimates allow only for limited and highly uncertain use in the planning of the electricity supply system. More reliable planning processes require considerably more information about future forecast quality.

[1] University of Kassel, Intelligent Embedded Systems, Wilhelmshöher Allee 73, 34121 Kassel, Germany
{j.schreiber, artjom.buschin, bsick}@uni-kassel.de

doi:10.18420/inf2019_74

Therefore, the article provides a comprehensive study on influences in forecast uncertainty, which has to be taken into consideration for future planning studies. The articles investigates uncertainty in four types of common ML models for wind and PV: Least absolute shrinkage and selection operator (LASSO), gradient boosting regression tree (GBRT), support vector regression (SVR), and multi-layer perceptron (MLP). Models are trained to forecasts the estimated day-ahead power generation based on NWP features as input. By repeated training with different test datasets, we create forecasts over the entire datasets for later analysis.

In the next step, we compare error distributions for each model concerning known influencing factors: Amount of training samples, forecast horizon, the terrain of wind farms, and a comparison of the uncertainties between the different machine learning (ML) models. By comparing binned forecasts errors, e.g., for different forecast horizons, with the Kullback-Leibler Divergence (KLD), we measure similarities and differences of these distributions. This comparison allows estimating when a bin is substantially different compared to a baseline and therefore gives insights to influential factors. Further, bins are compared with the Kruskal-Wallis [HS18] hypothesis test to verify a significant difference. The main contributions of this article are:

- We utilize common ML models, the grid search algorithm to find optimal model parameters, and common feature engineering techniques to provide forecasts results for wind and PV farms. By repeated training of the models on different training sets, we create forecasts for the complete dataset.
- For wind power forecasts, we show that the amount of training samples influences the forecast errors up to a certain threshold of data coverage (where data coverage is the proportion of the maximum number of data samples in the dataset to the actual amount of data samples within the historical data).
- Analyzing seasonal patterns reveal different influences for wind and PV that are related to the specific weather conditions within the individual seasons. Interestingly, forecast errors of adjacent seasons are not necessarily similar to each other in PV forecasts.
- The comparative study of ML forecasts models shows, that wind forecast errors within a similar terrain are more alike than for a similar amount of training samples. This relation suggests that forecasts models are more alike when external influences are excluded.

The remainder of this article is structured as follows. In Section 2 we detail related work. Section 3 outlines evaluation measures and applied ML models. Section 4 describes the experimental design and evaluation results w.r.t. data coverage, seasonal patterns, terrain, and model differences. Finally, we conclude our work and propose future work in Section 5.

2 Related Work

In current planning studies on future energy systems, considerations on the current and future uncertainty of power forecasts are only inconsiderably taken into account. The German Dena II study [KAS10], e.g., only considers forecasts error up to a horizon of two hours, neglecting that an increasing amount of renewable requires a larger forecast horizon such as day-ahead forecasts. Further, the study is missing an analysis of seasonal effects and forecast model specific uncertainties.

These (mostly) missing influential factors [Ya15] categorizes into the NWP input data, the power curve, and the prediction algorithm. The thesis of M. Lange [La03] relates forecasts error to the NWP data. In particular, the forecast uncertainty is assessed with respect to (w.r.t.) certain meteorological situations. However, the study only employs a physical model of the power curve with error correction and spatial refinement.

In the work of [Pi06], time series analysis techniques (e.g., ARIMA, ARX, Box-Jenkins) and a physical model are used to evaluate the forecasting skill. The author includes an analysis for different forecast horizons based on R^2 and root mean squared error (RMSE). Further, it contains a small subsection on the evaluation of the error distribution.

The results of [Mö04] include an analysis for time horizons between zero to nine hours and up to five days ahead. But the study is again focusing on the physical model and not considering machine learning models. Also [HMS13; KHP15] are focusing on time series analysis techniques (NARX) and (adapted) physical models, for uncertainty analysis.

More recently, in [Ge18] uncertainty in ML models, such as MLP, SVR, and an ensemble technique, are analyzed, but the thesis misses an evaluation for error distributions relating to different forecast horizons of wind power. Also, in [Re17] an analysis of ML models such as extreme gradient boosting technique, random forest, adaptive boosting, and persistence method is used to access the economic value for PV power generation.

[BT18] uses physical and semi-physical models for developing a forecast methodology for households that do not have access to solar irradiance information and are therefore limited to discrete weather information. The results are analyzed w.r.t. the weather features.

An interesting approach presents [NH18], in which the kriging method interpolate data with geographical properties for a location with no available data. A Naïve Bayes classifier along with a Gaussian probability distribution based on the overall data performs day-ahead forecasts of solar power based on the probability in one-hour intervals. The method is evaluated against the persistence model with mean absolute error (MAE) for different months of a year.

The simulation in [NR13] creates uncertainties for PV at different time scales to evaluate the economic and reliability effect for the grid. As it is a simulation tool, it is different from

ML models. The proposed method in [MOO18] allows for modeling the PV uncertainty based on past observations by using multivariate normal distributions.

The literature review shows that most of the work is focusing on models relating to time series analysis techniques and physical models. Further, the reviewed articles are missing a quantified comparison between the distribution of uncertainties or are even missing an in-depth analysis of the error distribution.

3 Method

This section gives a summary on common ML algorithms and presents their differences, to evaluate influences in forecast uncertainty. By using different ML algorithms, we assure to cover a broad spectrum in forecast errors. In the final section, we summarize error measures to estimate the deviation between actual and forecasted power generation.

3.1 Lasso

LASSO, also known as basis pursuit, is a linear model. Linear models typically provide a robust estimation, when NWPs are uncertain. Further linear models allow measuring the contribution of individual features through their coefficient, hence, making them highly relevant for analysis on error origin [HTF01]. In contrast to other linear regression models, LASSO allows for automatic selection of essential features. This selection is achieved by L_1 penalty, that effectively causes the coefficient of features to be exactly zero and hence excluding individual features.

3.2 Support Vector Regression

SVR is based on the concept of support vector machines (SVMs) for classification with changes in the definition of the optimization problem. One appealing property of SVMs is that the determination of parameters is locally and globally optimal due to the convex optimization [Bi06]. Further, by making use of the *kernel-trick* original NWP input features are transformed in a higher dimensional, even infinite dimensional, space. Transforming features into a higher-dimensional space provides features that are linearly separable [Va00]. The transformed features allow the SVR to achieve good results in many applications [Bi06], making them highly relevant for the evaluation of forecast uncertainty.

3.3 Multi-layer Perceptron

MLPs and more recently deep neural networks are a common technique for regression and classification tasks. In a feed forward MLP input features are transformed using

matrix multiplication and a subsequent (mostly) non-linear transformation. The former two operations are summarized as layers and successive applications of these layers, where the output of one layer is the input to the next layers, allows us to find a good representation of the data. In the final layer, the output layer, a simple linear combination can be used for renewable energy forecast. Primarily through their capability to find suitable representations of the NWP data, MLPs achieve state of the art performance in renewable power forecast [Ge16]. This performance makes them highly relevant for the evaluation of forecast uncertainty.

3.4 Gradient-Boosting-Regression-Tree

GBRT originate from the idea, that a combination of weak learners improves the overall performance. Therefore, the gradient boosting algorithm trains trees in regions of most substantial forecast error. The ensemble technique combines the individual trees improving the overall performance. A single tree partitions the features space in a set of rectangles and estimates a constant forecast value for each rectangle [HTF01]. The partitioning provides an interpretable structure to explain forecast decision, which is not feasible with SVR and MLPs. Further, the algorithm is not making use of any data representation techniques as with these approaches.

3.5 Error Measures

To assess influences in forecast uncertainty through forecast errors, it is essential to evaluate the error with $e = y - \hat{y}$. It gives insights between the actual power generation y compared to the forecasted power $\hat{y}_i$. In contrast to mean based measures, e provides the most detailed view on the error; combined with a visualization of the error distribution through a histogram or a boxplot it allows to assess skewness and other statistical measures of the error distribution. A comprehensive analysis of deterministic error measures in the field of renewable energy forecast is given in [Ge18]. The results can be summarized as follows

- The coefficient of determination R^2 assesses how much of the variance in the historical power data is explained by the model. As it is only capable of evaluating the amount of linear correlation, it is often used as a measure to compare different forecast techniques.
- To account for extreme errors of e, quadratic errors such as the mean squared error (MSE) are recommend.
- Absolute measures such as MAE are suited for monetary evaluation criteria (linear evaluation criteria).

In the following, we will stick to e^2 as it allows for comparison of overall forecast quality of the model by terms of mean (MSE) and median, especially when visualized via boxplot.

To compare distributions of errors with each other, we use the KLD. The KLD is a non-symmetric statistical measurement to determine the difference between two distributions allowing to quantify the similarity, e. g., between the error distribution from the GBRT and the SVR.

4 Experimental Evaluation

In the following, we provide analysis on error distributions from different ML models and measure their similarity to another for wind and PV. By estimating the KLD between distributions for different (external) factors, we get insights on how they relate to each other. Therefore, we first give details on the model training and the two datasets. The first study estimates influences caused by a limited amount of training samples for wind power forecasts. Results are evaluated w.r.t. the data coverage, where data coverage refers to the proportion between the maximum number of data samples to the actual amount of data samples within the dataset. In the next section, we analyze seasonal influences such as the hour of the day or season of the year for wind and PV as well as terrain specific influences in the wind dataset. As results for PV models suggest that there are strong seasonal patterns to consider - that are less present for wind models - we limit the final analysis to the WindFarm dataset. Limiting these and other external influences allow to compare the error distributions of the different power forecasting models.

4.1 Design of Experiment

For the following two datasets we train the LASSO, SVR, MLP, and GBRT to forecast the power generation optimized through grid-search for the following parameters[2][3]:

- **LASSO**: Alpha (0.1 - 1) and maximum number of iterations (1000-4000).
- **SVR**: Penalty (0.01-0.001) and gamma (0.01-1000).
- **GBRT**: Max depth (12 -20), minimum samples split (2-128), number of estimators (1200-2000), and learning rate (0.05-0.2).
- **MLP**: Maximum number of iterations (10000), early stopping (true or false), learning rate ($0.01\text{-}1 \times 10^{-6}$), learning rate type (constant or adaptive), and hidden layer sizes ((100, 50, 30, 15, 5), (50, 100, 200, 50), (20, 10, 5)).

Solar Farm Dataset: The *SolarFarm* dataset consists of 114 PV facilities in Germany. Their installed nominal power ranges between 7.2kW and 12573kW. The dataset has a

[2] Other model parameters are the default parameters by scikit-learn, last-accessed June 2015.

[3] Parameters are selected to obtain a similar training time for all models (except LASSO as this is a linear model) while achieving a good performance based on initial testing.

three-hour resolution and is recorded from the beginning of 2016 to the end of April 2017, resulting in a maximum of 3880 data points. In total the dataset has 51 input features as input. Features with correlation to the power generation (e.g., sun position, solar height, clear sky, and radiation) are shifted in time by three hours to take future and past effects of the weather into account for prediction.

Wind Farm Dataset: The *WindFarm* dataset contains the power generation taken from 54 wind farms that are distributed throughout Germany. These values were recorded hourly over two years (2016 and 2017), resulting in a maximum of 17520 data points. The dataset contains information about the terrain of each farm (flatland, forest, and offshore). In total the dataset has 7 NWP features as input. Features of wind speed and wind direction influencing the power generation [SS18] are time-shifted by two hours to take future and past effects of the weather into account for prediction.

Both datasets were manually filtered to remove outliers, e.g., caused by maintenance. Depending on the number of outliers and maintenance the amount of data coverage ranges between 50 and 100 percent for wind data w.r.t. recorded period, where data coverage refers to the proportion between the maximum number of data samples to the actual amount of data samples within the dataset. The data coverage for PV is mostly above 90%.

To compare forecast errors, we normalize the generated power by the maximum power generation. Input features are standardized for zero mean and unit variance based on the training dataset in each run. We optimize each model through a grid search on the validation dataset. To make the best use of the full data range, we use different runs of the experiment to shift the test data throughout the recorded period: Six months for the wind and four months for the PV dataset resulting in four runs for each dataset. In each run, the remaining data is used for training (80%) and validation (20%). After completing all training runs, combing predictions from all test datasets provides an *evaluation dataset* for estimating influences in the entire period of the original data. To account for extreme errors and measure the quality between a single forecast and the historical power, we use the squared error. We fit distributions of the squared error with the $\tilde{\chi}^2$ distribution to compare them with the KLD

4.2 Influence of the Amount of Training Data

The digitalization of the current and future energy market will provide an increasing amount of training data. To determine the extent to which the amount of training data influences the forecast error, we analyze it in this section.

Therefore we estimate the data coverage of a farm in percent compared to the maximum number of data points. It turns out that the data coverage in PV farms is consistently above 92% except for one farm, respectively, we do not consider it in further analysis. However, the data coverage from wind farm range between 49 and 99% allowing for clustering them in ten percent steps, see Figure 1. As the size of the test dataset is constant in each run, the

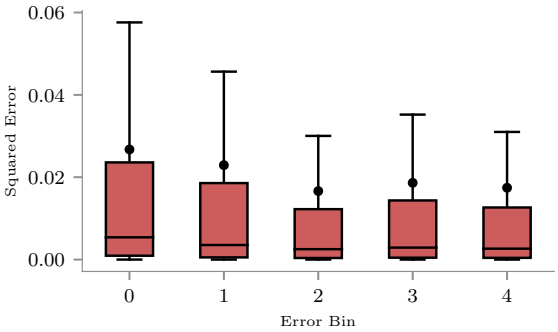


Fig. 1: Boxplot on squared error for bins of data coverage for the WindFarm dataset based on the MLP model. From left to right the data coverage is as follows: 50 – 60%, 60 – 70%, 70 – 80%, 80 – 90%, and 90 – 100%. The mean of the error is visualized as a dot.

Bins	0	1	2	3	4
0	0.000	0.003	0.044	0.037	0.065
1		0.000	0.068	0.059	0.094
2			0.000	0.000	0.002
3				0.000	0.004

Tab. 1: KLD measuring the similarity between different amounts of data coverage for the error of the GBRT model on the WindFarm dataset. From one to five the data coverage is as follows: 50 – 60%, 60 – 70%, 70 – 80%, 80 – 90%, and 90 – 100%.

size of the training data is as well, respectively, the data coverage is directly linked to the amount of training data and will be treated equally in the following.

Figure 1 shows the relation between the number of training samples and the error: With the increasing amount of data, the median as well as the mean decrease. The spread of the error is similar for bin two, three, and four. Bin zero and one have a broader spread of the error. The decreasing mean, median, and spread show that there is a relation between the amount of data for training and the forecast error.

To verify a significant difference between these bins, we compare them with the KLD and the Kruskal-Wallis hypothesis test. Kruskal-Wallis hypothesis shows that the forecast error for all ML models and all bins of data coverage are significantly different at a significance level of $\alpha = 5\%$. The exemplary results in Table 1, highlight the previous observation: A decreasing data coverage, causes an increased spread, median, and mean resulting in larger values of the KLD, e.g., when comparing bin zero with bin four. Bin two, three, and four are quite similar to each other nonetheless.

As expected, there is a relation between the amount of data available and the forecast error.

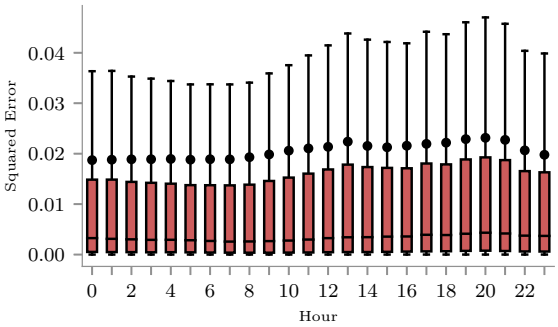


Fig. 2: Boxplot on squared error for bins of the hour of the day for the WindFarm dataset based on the MLP model. The mean of the error is visualized as a dot.

With an increasing amount of available data, the ML model tends towards a minimum error, the NWP input data probably cause that.

4.3 Influences by Seasonal Patterns and Terrain

Seasonal influences that are present in seasons of a year or hours of a day are well known. Nonetheless, there is limited research on how these patterns affect the error distribution in wind and PV forecast based on ML techniques. More common is the analysis of forecast error w.r.t. their terrain, which this section also covers.

In the following, we address season of a year and the hour of the day. In terms of PV, the hour of the day has two meanings. First, due to the daily pattern of the sun, we can observe patterns within the power generation. Second, with the rising time of the day, the forecast horizon of the NWP model increases (as the so-called NWP model run typically originates from 12 UTC). As the horizon increases, the error of the weather forecast increases and respectively that of the power forecast model. The latter also holds for wind power forecasts.

In the sample boxplot, Figure 2 for wind errors, we can observe this pattern. The median and mean errors for different hours of the day do not increase drastically due to the absence of seasonal weather patterns in the wind; detailed observations exist when measuring similarity through KLD in Table 2. The errors at the end of the day are more similar to each other employing the KLD, compared to the origin of the weather forecast due to the increased forecast error of the NWPs. Nonetheless, all errors, when comparing different hours of the day, are significantly different in the Kruskal-Wallis hypothesis test ($\alpha = 5\%$) except: Four cases in the linear model, one for the MLP, and two comparisons for the GBRT model.

Hour	0	3	6	9	12	15	18	21
0	0.000	0.000	0.057	1.905	7.619	5.150	0.831	0.437
3		0.000	0.067	1.975	7.815	5.294	0.873	0.467
6			0.000	1.212	5.636	3.707	0.431	0.173
9				0.000	1.003	0.449	0.168	0.423
12					0.000	0.094	2.190	3.168
15						0.000	1.244	1.932
18							0.000	0.054

Tab. 2: KLD measuring the similarity between hours of a day for the error of the MLP model on the PVFarm dataset.

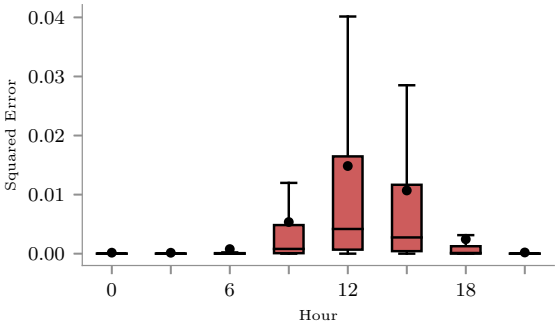


Fig. 3: Boxplot on squared error for bins of an hour of the day for the SolarFarm dataset based on the GBRT model. The mean of the error is visualized as a dot.

For PV however, see Figure 3, we can observe a strong seasonal pattern in the error distributions for different hours of the day. This observation is also present when estimating the KLD, resulting in large values when comparing 12 with 0 o'clock, see Table 2. This seasonal pattern is to expected, as during the night there is no power generation and respectively the difference in the error distribution is notable when compared to the day. Compared to wind, there are also more considerable differences in the error distributions during the daytime. The daily pattern of the sun causes these differences that result in different error distributions. Again, all errors, when comparing different hours of the day, are significantly different in the Kruskal-Wallis hypothesis test ($\alpha = 5\%$) except: Four cases in the SVR model and one case for the MLP model.

In the analysis for different seasons of a year for wind, we observe that in the third season all models and datasets have the lowest median, mean, and spread of the error for wind. In other seasons of the year, extreme weather conditions are more common, causing larger error values. The KLD, see Table 3, confirms our intuition, that error distributions for seasons close to each other are more similar to another than those far away.

Season	1	2	3	4
1	0.00	0.02	0.19	0.10
2		0.00	0.09	0.04
3			0.00	0.01

Tab. 3: KLD measuring the similarity between seasons of a year for the error of the SVR model on the WindFarm dataset. Season one equals winter, season two equals spring, etc..

Season	1	2	3	4
1	0.00	0.18	3.21	0.08
2		0.00	1.58	0.02
3			0.00	2.02

Tab. 4: KLD measuring the similarity between seasons of a year for the error of the SVR model on the SolarFarm dataset. Season one equals winter, season two equals spring, etc..

Contrary to wind forecasts, PV models have more substantial errors in the third season. In other seasons of the year, the different position of the sun causes a different amount of direct and diffuse radiation making it the forecast model easier to forecast the power generation. For instance, the solar radiation (direct and diffuse) is the smallest in season one in the dataset. The analysis of the KLD in Table 4 suggest that the difference of uncertainty is significant even for seasons of the year that are close to another. These differences are caused by the larger magnitude of the forecast error, especially in the third season. Only when comparing season one and three for errors of the GBRT model on the WindFarm dataset the Kruskal-Wallis ($\alpha = 0.05$) estimates no significant difference.

The analysis of the terrain in Figure 4 shows that the smallest errors are present for parks located in a farmland terrain. All errors, when comparing the different terrains, are significantly different in the Kruskal-Wallis hypothesis test ($\alpha = 5\%$), where farmland has the smallest, forest the second smallest, and offshore the most substantial forecast error. Note that the terrains have a varying amount of farms. Farmland has 37, the forest has 11, and offshore includes four farms. Even though the terrains have a varying amount of farms, the larger error for offshore parks might be caused by large ramping events of the wind. Interestingly, when measuring the similarity, the error distribution of offshore farms is closer to the farmland, than farmland to the forest employing the KLD. This smaller KLD might be due to more complex weather conditions in the forest and offshore terrain. For instance, turbulence on the sea might be similarly present in forests (that are often also elevated) causing a similar uncertainty distribution.

Conclusively, we showed similarity and dissimilarity in seasonal and terrain specific patterns. Interestingly, the difference in error distribution is one of the largest for different seasons of the year for wind and PV. Wind errors are more significant in the winter and autumn, while PV models have larger values in the spring and summer time. Finally, the uncertainty distribution in offshore terrain is like that of the forest.

4.4 Influences by Models

After analyzing external influences to the error distributions, in this section, we are interested in comparing the similarity between the ML models. As results for PV models suggest that

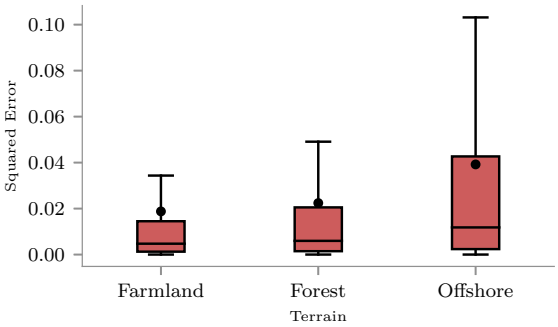


Fig. 4: Boxplot on squared error for different terrains for the WindFarm dataset based on the GBRT model. The mean of the error is visualized as a dot.

	GBRT	LASSO	SVR	MLP
GBRT	0.00	1.53	0.05	0.14
LASSO		0.00	0.98	0.66
SVR			0.00	0.02

Tab. 5: KLD measuring the similarity between ML models with data coverage between 90 and 100%.

	GBRT	LASSO	SVR	MLP
GBRT	0.00	1.46	0.05	0.15
LASSO		0.00	0.94	0.61
SVR			0.00	0.03

Tab. 6: KLD measuring the similarity between ML models for a farmland terrain.

there are strong seasonal patterns to consider - that are less present for wind models - we limit further analysis to the WindFarm dataset.

In previous results from wind models, we show that the error is the smallest for the farmland terrain and when training the ML model on a data coverage between 90 to 100%.

Limiting the analysis to the smallest errors gives us insights, more similar to an absence of external influences, as those influences must have a smaller effect on the error distribution compared to distribution with larger mean, median, and spread. Ultimately, allowing to access the differences in the ML models and not those caused by external influences.

In both analyses, we observe that GBRT achieves the smallest error, SVR the second, MLP the third lowest and LASSO has the most substantial error. Table 5 and 6 summarize the the difference in their error distributions for the experiment with maximum data coverage and the farmland terrain.

Results suggest that distributions within a terrain are more similar to each other than within maximum data coverage caused by the relation that specific weather conditions, are individual for different terrains, resulting in terrain specific forecast errors. Nonetheless,

estimates of the Kruskal-Wallis hypothesis test ($\alpha = 5\%$) shows that they are still substantially different.

5 Conclusion and Future Work

In this article, we presented an in-depth analysis and comparison for influencing factors of uncertainty in wind and PV power forecasts based on four different ML models. In our analysis, we found substantial influences and differences between compared bins of uncertainty that reveal the need to consider them in future planning studies.

For instance, the study reveals strong seasonal patterns in the uncertainty for wind and PV. For wind power forecasts, neighboring seasons and hours are similar to each other. For seasonal patterns within a year, these forecasts will benefit from optimizing NWP forecasts for extreme weather situations that cause substantial errors in the winter time. Due to significantly larger errors for the third season adjacent seasonal bins in PV forecasts are not necessarily similar to each other. Similar results are obtained for daily patterns. For daily patterns, we recommend to use NWP forecasts that are closer to the time (noon) of most substantial error.

By analyzing the relation between the amount of training data and the uncertainty, we showed that models improve when using additional data up to a data coverage of about 70%. Reducing this error further is, e.g., possible with deep learning models that have a higher capacity to learn the relation between NWP features, e.g., they can be trained across several parks in a multi-task-learning setting. However, even with an increasing amount of data, the minimum forecast error will be limited to that error caused by the NWP.

The study reveals that after minimizing external influences, differences in the uncertainty distributions from the four ML models are still present motivating the need to consider the underlying forecast model in future planning studies.

In the future, we aim to repeat the analysis for intra-day, different NWP models, and data with a 15-minute resolution. Further, we will investigate how *transfer learning* can be utilized to reduce forecast uncertainty when limited data is available.

Acknowledgement This work was supported within the project Prophesy (0324104A) funded by BMWi (Deusches Bundesministerium für Wirtschaft und Energie / German Federal Ministry for Economic Affairs and Energy).

References

[Bi06] Bishop, C. M.: Pattern Recognition and Machine Learning. Springer, 2006.

[BT18] Brecl, K.; Topic, M.: Photovoltaics (PV) system energy forecast on the basis of the local weather forecast: Problems, uncertainties and solutions. Energies 11/5, p. 12, 2018.

[Ge16] Gensler, A.; Henze, J.; Sick, B.; Raabe, N.: Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks. In: SMC. IEEE, pp. 002858–002865, 2016.

[Ge18] Gensler, A.: Wind power ensemble forecasting, PhD thesis, Univerisity of Kasel, 2018, p. 204.

[HMS13] Holttinen, H.; Miettinen, J.; Sillanpää, S.: Wind power forecasting accuracy and uncertainty in Finland. 2013.

[HS18] Hedderich, J.; Sachs, L. 1.-.: Angewandte Statistik Methodensammlung mit R. Springer Spektrum, 2018.

[HTF01] Hastie, T.; Tibshirani, R.; Friedman, J.: The Elements of Statistical Learning. Springer, 2001.

[KAS10] Kohler, S.; Agricola, A.-C.; Seidl, H.: Integration erneuerbarer Energien in die deutsche Stromversorgung im Zeitraum 2015-2020 mit Ausblick 2025, tech. rep., 2010, p. 564.

[KHP15] Ko, W.; Hur, D.; Park, J.-K.: Correction of wind power forecasting by considering wind speed forecast error. JICEE 5/1, pp. 47–50, 2015.

[La03] Lange, M.: Analysis of the Uncertainty of Wind Power Predictions, PhD thesis, 2003, p. 139.

[Mö04] Möhrlen, C.: Uncertainty in wind energy forecasting, PhD thesis, University College Cork, 2004, p. 178.

[MOO18] Murata, A.; Ohtake, H.; Oozeki, T.: Modeling of uncertainty of solar irradiance forecasts on numerical weather predictions with the estimation of multiple confidence intervals. Renewable Energy 117/, pp. 193–201, 2018.

[NH18] Nam, S.; Hur, J.: Probabilistic Forecasting Model of Solar Power Outputs Based on the Naive Bayes Classifier and Kriging Models. Energies 11/11, p. 15, 2018.

[NR13] NREL: Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales, tech. rep. May, 2013, p. 41.

[Pi06] Pinson, P.: Estimation of the uncertainty, PhD thesis, Ecole des Mines de Paris, 2006, p. 266.

[Re17] Reindl, T.; Walsh, W.; Yanqin, Z.; Bieri, M.: Energy meteorology for accurate forecasting of PV power output on different time horizons. In: Energy Procedia. Vol. 130, pp. 130–138, 2017.

[SS18] Schreiber, J.; Sick, B.: Quantifying the Influences on Probabilistic Wind Power Forecasts. In: ICPRE. Vol. 3, p. 6, 2018.

[Va00] Vapnik, V. N.: The Nature of Statistical Learning Theory. Springer-Verlag, 2000, ISBN: 978-0-387-98780-4.

[Ya15] Yan, J.; Liu, Y.; Han, S.; Wang, Y.; Feng, S.: Reviews on uncertainty analysis of wind power forecasting. Renewable and Sustainable Energy Reviews 52/, pp. 1322–1330, 2015.

Track 7 – Digitale Bildung

Digitale Bildung

Nadine Bergner[1], Ira Diethelm[2]

Bildung in der digitalen Welt erfordert sowohl informatische Bildung als auch Medienbildung und muss alle Kinder, Jugendlichen und auch Erwachsenen erreichen. Nur so können sie unsere von der Digitalisierung durchdrungene Welt reflektiert und aktiv mitgestalten. Sowohl die allgemeinbildende Schule als auch außerschulische Lernorte und berufliche Aus- und Weiterbildungsprogramme bis zu privaten Fortbildungen wollen und müssen ihren Beitrag dazu leisten. Wie digital(isiert) die Welt um uns herum bereits ist, wird täglich spürbar – im Bildungssystem ist sie jedoch als Unterrichtsgegenstand vielerorts noch nicht angekommen. In diesem Track finden Projektideen, Erfahrungsberichte, Evaluationen und Aufbereitungen von Theorie Platz, welche dazu beitragen Digitale Bildung mit dem Ziel der Mündigkeit in der digitalen Welt in der Breite zu fördern.

In diesem Track werden in drei thematischen Sessions insgesamt neun Beiträge präsentiert und diskutiert. Die Sessions beschäftigen sich dabei mit folgenden Schwerpunkten:

Session 1: Grundlagen der Digitalen Bildung

Session 2: Digitale Bildung in spezifischen Kontexten

Session 3: Digitale Bildung von Seniorinnen und Senioren

Die Beiträge einschließlich eines Extended Abstracts sind im Folgenden in eben dieser Reihung abgedruckt.

Die erste Session beschäftigt sich mit den (theoretischen) Grundlagen der Digitalen Bildung und umfasst folgende Beiträge: Critical Computational Thinking: Konzeptentwurf zur Vermittlung Informatikwissen für die Digitalisierungsgestaltung (Esther Ruiz Ben), Informatik für alle - Eine Analyse von Argumenten und Argumentationsschemata für das Schulfach Informatik (Stefan Seegerer, Tilman Michaeli and Ralf Romeike) sowie IT-Sicherheit und Medienkompetenz – Digitale Bildung und Mündigkeit im Kontext der Wissenskluft im tertiären Bildungsbereich (Raphael Morisco).

Die zweite Session beschäftigt sich mit der Vermittlung digitaler Bildung in drei spezifischen Anwendungskontexten: Informatische Bildung als Verbraucherschutz für reflektierte Handlungen in der digitalen Welt (Manuel Froitzheim, Michael Schuhen and Timo Stentenbach),

[1] TU Dresden, nadine.bergner@tu-dresden.de
[2] Carl von Ossietzky Universität Oldenburg, ira.diethelm@uni-oldenburg.de

doi:10.18420/inf2019_75

Erfassung von Medienkompetenz innerhalb von E-Learning-Systemen am Beispiel der Meisterausbildung im Stuckateur-Handwerk (Kim Petry, Tobias Greff and Dirkt Werth) und Chancen und Herausforderungen von Virtual Reality in der Aus- und Weiterbildung im Gesundheitswesen (Julian Schuir, Alina Behne and Frank Teuteberg).

Die dritte Session umfasst drei Beiträge, die explizit die Zielgruppe Seniorinnen und Senioren adressieren: Technikbegleitung. Aufbau von Initiativen zur Stärkung der Teilhabe Älterer im Quartier (Elisabeth Bubolz-Lutz and Janina Stiel), Zu alt für Informatik?: Seniorinnen und Senioren erobern die digitale Welt (Svenja Noichl and Ulrik Schroeder) und Development of a senior-friendly training concept for imparting media literacy (Sebastian Wilhelm, Dietmar Jakob and Melanie Dietmeier).

Die Beiträge der Sessions wurden im Reviewprozess aus insgesamt 16 Einreichungen (15 Vollbeiträge und ein Extended Abstract) für den Track Digitale Bildung ausgewählt. Die angenommenen Beiträge sind sehr vielfältig und decken so auch thematisch die Spannweite der eingereichten Beiträge ab. Die Auswahl erfolgte nach gängigen Qualitätskriterien für GI-Konferenzen dieses Bereichs. Die Themen der Sessions ergaben sich aus der Gruppierung der angenommenen Beiträge.

Im Track-Programmkomitee haben neben uns hierzu folgende Personen mitgewirkt, denen an dieser Stelle unserer besonderer Dank gilt:

- Torsten Brinda, Universität Duisburg-Essen
- Michael Fothe, Universität Jena
- Jens Gallenbacher, TU Darmstadt
- Lutz Hellmig, Universität Rostock
- Thomas Knaus, FTzM | PH Ludwigsburg
- Torsten Otto, Universität Hamburg
- Ralf Romeike, FU Berlin
- Ulrik Schroeder, RWTH Aachen
- Carsten Schulte, Universität Paderborn

Full Papers

Critical Computational Thinking: Konzeptentwurf zur Vermittlung von Informatikwissen für die Digitalisierungsgestaltung

Esther Ruiz Ben[1]

Abstract: Digitalisierungsprozesse beeinflussen die Zusammenarbeit zwischen verschiedenen Disziplinen und Arbeitsfeldern. Informatikbezogene Kenntnisse und Fertigkeiten werden einerseits zunehmend in unterschiedlichen Tätigkeitsfeldern verlangt, andererseits wird von Informatikspezialist*innen erwartet, dass sie in der Lage sind, ihre Informatikkenntnisse flexibel auf die unterschiedlichsten Anwendungsgebiete zu übertragen. Gleichzeitig wird durch Digitalisierungsprozesse die Trennung zwischen Privatem und Beruflichem immer undeutlicher, so dass digitale Kompetenzen in allen Lebensbereichen immer wichtiger werden. Doch die interdisziplinäre Zusammenarbeit, die in der beruflichen und forschungsbezogenen Praxis verlangt wird, ist in der Wissensvermittlung, sowohl in der universitären als auch in der schulischen Lehre, noch nicht etabliert. Die Vermittlung digitaler Kompetenzen wird ebenfalls selten in die Lehre integriert. Informatik wird oft als Programmierung zusammengefasst und digitale Kompetenzen als Nutzungskompetenz in Bezug auf das Internet beschränkt. Um einen interdisziplinären Dialog zwischen bisher getrennten Disziplinen bzw. Fächern zu entfalten und damit die Vermittlung von Informatikwissen mit digitalen Kompetenzen zu vereinbaren, die jenseits der Nutzung auch die Gestaltung von digitalen Technologien reflektieren, schlage ich die Umsetzung eines „Critical Computational Thinking" vor. „Critical Computational Thinking" (CCT) kombiniert Computational Thinking-Prinzipien mit ethischen Komponenten sowohl der Gestaltung, Anwendung als auch der Nutzung von Digitalisierungsprozessen und –produkten, die im Konzept der digitalen Kompetenzen berücksichtigt werden. Im vorliegenden Beitrag skizziere ich dieses Konzept, das aktuell in einem BMBF Projekt angewendet wird, und diskutiere, welche konkreten Aspekte beider Ansätze im Zusammenhang mit Digitalisierungsprozessen für die Informatiklehre relevant sind.

Keywords: Digitalisierung, Computational Thinking, Ethik, Digitale Kompetenzen, Informatikdidaktik.

1 Einführung

Digitalisierungsprozesse stellen alle Lebensbereiche vor große Herausforderungen. In der Arbeitswelt müssen heutzutage Berufe, die früher getrennt operiert haben, nicht nur miteinander arbeiten, sondern darüber hinaus auch vermehrt mit Informatikspezialist*innen, um den Anforderungen der Digitalisierung gerecht zu werden. Computerbezogene Kenntnisse und Fertigkeiten werden immer mehr von Arbeitnehmer*innen verlangt, während von Informatiker*innen erwartet wird, dass sie ihre Informatikkenntnisse flexibel in die unterschiedlichsten Anwendungsgebiete übertragen können. Auch in der

[1] Email: esther.ruiz-ben@campus.tu-berlin.de

doi:10.18420/inf2019_76

akademischen Forschung werden interdisziplinäre Projekte zunehmend gefördert. Doch in der Wissensvermittlung der Informatik, sowohl in der universitären als auch in der schulischen Lehre, wird Interdisziplinarität noch wenig praktiziert. Die Diskrepanz zwischen den beruflichen Anforderungen und der Vermittlung von Informatikwissen ist vor allem in schulischen Bereichen bemerkbar. Parallel dazu werden die zunehmenden Anforderungen an digitale Kompetenzen im Alltag (z. B. im Zusammenhang mit hate speech) in schulischen Kontexten wenig beachtet oder in restriktiver Form in einzelnen Fachbereichen und nicht fachübergreifend thematisiert. Grundsätzlich wird die Wissensvermittlung der Informatik mit dem Erlernen von Programmierung und Problemlösungsstrategien gleichgesetzt. Die Reflexion darüber, wie solche gesellschaftlichen „Probleme" überhaupt definiert werden bzw. über den Sinn und die Rolle der Informatik in Digitalisierungsprozessen, bleibt weitestgehend aus. Fragen über die Rolle der Informatik bzw. von Informatiker*innen in Zusammenarbeit mit anderen Fachbereichen bei der Gestaltung von Digitalisierungsprozessen oder bei der Reflexion der Implikationen von Digitalisierung werden nicht thematisiert. Um einen interdisziplinären Dialog zwischen bisher getrennten Disziplinen bzw. Fächern zu entfalten und damit die Vermittlung von Informatikwissen für die gemeinsame Gestaltung von Digitalisierungsprozessen zu etablieren, schlage ich die Umsetzung eines „Critical Computational Thinking" sowohl in der schulischen Wissensvermittlung als auch an der Hochschule vor.

„Critical Computational Thinking" (CCT) konzeptualisiere ich als eine interdisziplinäre Kombination aus gemeinsamen wissenschaftlichen Ansätzen und Prinzipien, die im Computational Thinking betont werden, und ethischen Komponenten[2] [Mi03] der Gestaltung, Anwendung und Nutzung von Digitalisierungsprozessen und –produkten, die im Konzept der digitalen Kompetenzen berücksichtigt werden und zum Teil in der Informatik und in der Forschung über die Professionalisierung der Informatik bereits diskutiert werden. Das bedeutet, dass, ergänzend zu den Kritiken am Konzept des Computational Thinking [TD16], eine ethische Kontextualisierung und Auseinandersetzung in der Gestaltung von digitalen Produkten und Dienstleistungen in die Vermittlung von Informatikmethoden integriert wird. Dazu gehören Ansätze, die auf eine offene bzw. inklusive Teilhabe an der Gestaltung von digitalen Produkten und Dienstleistungen – nicht nur an ihren Konsequenzen - abzielen und die bereits existierende Strategien (z. B. GI Manifesto, KMK-Strategie, Bildungsstandards, etc.) ergänzen. Informatikwissen wird damit nicht auf Methodenentwicklung bzw. –umsetzung und „Problemlösung" reduziert, sondern um Aspekte einer Professionalisierung der Informatikpraxis erweitert. Diese benötigen zunächst eine kontextualisierte Problemdiagnose in Digitalisierungsprozessen, bevor die Problemlösung gesucht wird. Das bedeutet, dass die Vermittlung von Informatikwissen Komponenten der Professionalität (Diagnose, Inferenz, Behandlung [Ab88]) integrieren sollte. Dies müsste bereits in der schulischen Lehre und darauf aufbauend in den unterschiedlichen Feldern der beruflichen Informatikbildung geschehen.

[2] Critical thinking ist ein Bildungsziel [De10], das aus sehr unterschiedlichen philosophischen Traditionen definiert wurde. Einige Dimensionen aus diesem theoretischen Bildungsziel wie zum Beispiel „inferring", "judging" oder „deciding" sind ähnlich zu den Professionalitätsaspekten des hier skizzierten Konzeptes des Critical Computational Thinking. So definiert beispielsweise auch Facione {Fa90:25] „internal dispositions" as *„those that contribute causally to doing a good job of thinking critically once one has started"*.

Kontextualisierende Fragen, wie zum Beispiel: „Was ist wünschenswert? Für wen ist das wünschenswert?“ [Gr18] Welche und wessen „Probleme“ in Digitalisierungsprozessen werden als relevant betrachtet und warum? Welche Informatikmethoden bzw. digitalen Tools werden als passend für die Problemlösung erachtet und welche möglichen, zusätzlichen „Probleme“ treten bei der Anwendung dieser Methoden auf? Diese Fragen gehören zum Konzept des Critical Computational Thinking. Sie können nur interdisziplinär beantwortet werden und erfordern nicht nur eine enge Verzahnung der Informatik mit anderen Fachbereichen, sondern auch eine Erweiterung der Definition davon, was Informatik ist, über das reine Programmieren hinaus. Neu an diesem Konzept für die Informatikdidaktik ist vor allem diese Betonung der Interdisziplinarität und der ethischen Kontextualisierung von informatischer Arbeit in konkreten Digitalisierungsprojekten. Das Setting für die Umsetzung von Critical Computational Thinking ist nicht auf schulische Kontexte begrenzt. Das Konzept kann auch in weiteren Settings der Vermittlung von Informatikwissen im Rahmen von Digitalisierungsprozessen umgesetzt/angewendet/auf...übertragen werden?

In diesem Paper skizziere ich das Konzept des „Critical Computational Thinking“, welches es ermöglicht, informatische Wissensvermittlung ethisch zu kontextualisieren. Meine Argumentation habe ich in zwei Teile gegliedert. Im ersten Teil fokussiere ich mich auf das Konzept des Computational Thinking und die Kritiken daran sowie auf die Vermittlung von digitalen Kompetenzen, um eine kritische Rekonzeptualisierung in diesen Diskursen zu positionieren und zu ergänzen. Im zweiten Teil erläutere ich die konkreten Aspekte von „Critical Computational Thinking“, die für die Vermittlung von Informatikwissen für die interdisziplinäre Gestaltung von Digitalisierungsprozessen relevant sind.

2 Computational Thinking und digitale Kompetenzen

Das Konzept einer Informatikdidaktik für die Schulen wurde bereits vor vierzig Jahren von Papert [Pa80] thematisiert. Jeannette Wing hat 2006 einige Aspekte dieses Konzepts übernommen und mit anderen didaktischen Prinzipien als Computational Thinking zusammengefasst und popularisiert. In den letzten zehn Jahren wurde Computational Thinking in die schulischen Curricula mehrerer Länder integriert. In der Fach-Literatur (i. e. [TD16]) wird jedoch kritisiert, dass soziale Aspekte bei der curricularen Umsetzung der Prinzipien wenig thematisiert werden. Der Fokus von Computational Thinking wird hauptsächlich auf das Erlernen von technisch zentrierten Informatikkenntnissen gelegt.

Wing schlägt vor, dass durch Computational Thinking Kompetenzen vermittelt werden sollen, die allen Menschen offenstehen. Fünf Kernthemen und –kompetenzen werden im Konzept des Computational Thinking konkretisiert: Probleme formulieren bzw. -dekomponieren, Algorithmen bilden, Muster erkennen und abstrahieren, Problemlösungen evaluieren [Wi06]. Diese grundsätzlichen Kompetenzen, die mathematische und ingenieurwissenschaftliche Prinzipien berühren, erachtet Wing als genauso wichtig wie

Schreiben, Lesen oder Arithmetik. Weitere Themen wie analytisches Denken für Problemlösungen, Systemdesign und das Verständnis menschlichen Verhaltens wurden später bei Wing (2008) in das Konzept des Computational Thinking integriert.

Wings Konzept wurde von Bildungsverbänden in den USA (z. B. Computer Science Teachers Association (CSTA)) übernommen bzw. adaptiert. Darüber hinaus wurde das Konzept von internationalen Bildungsverbänden, aber auch von großen IT Firmen weiterentwickelt. Daraus resultierten verschiedene Definitionen von Computational Thinking, die ursprünglich Wings Konzept nutzten. So zum Beispiel entwickelte The International Society for Technology in Education (ISTE) zusammen mit der CSTA eine Redefinition von Computational Thinking. In dieser Redefinition wird Computational Thinking als Prozess des Problemlösens konzeptualisiert und in sechs Grundkompetenzen zusammengefasst:

- Problemformulierung für eine unterstützende Nutzung von Computern für Lehrkräfte
- Logische Organisation und Analyse von Daten
- Daten in Modellen und Simulationen darstellen
- Automatisierung von Lösungen mit algorithmischen Prinzipien
- Identifikation, Analyse und Implementation der effizientesten und effektivsten Problemlösung mit einer Kombination von Schritten und Ressourcen
- Generalisierung und Transfer von Problemlösungsprozessen in andere Disziplinen

Die interdisziplinäre Dimension des Methodentransfers ist, ähnlich wie bei der Adaptierung des Konzepts, bei großen IT Firmen vorgesehen. So zum Beispiel fokussiert sich Google [Go11] auf vier Prinzipien: Zerlegung, Mustererkennung, Abstraktion und Algorithmen bilden. Es erklärt mit Beispielen, wie diese in verschiedene Fachbereiche übertragen werden können[3]. Gleichzeitig wird dennoch in diesen Redefinitionen und Adaptierungen Computational Thinking als eine Methode des Problemlösens mit informatischen Techniken konzeptualisiert. Speziell diese Charakterisierung von Computational Thinking als Problemlösungsmethode wurde in mehreren Ländern als Hintergrundgedanke für die Integration des Konzeptes in die schulischen Curricula genutzt. Legitimiert wurde die Einführung von Computational Thinking in Curricula mit der Idee, dass die Digitalisierung der Gesellschaft das Erlernen von informatischen Methoden erfordere [BHC11]. Darüber hinaus haben andere Autor*innen betont, dass die Integration von Computational Thinking in schulischen Curricula auch dazu beitragen könne, dass andere nicht-informatische Fächer besser von Informatiker*innen verstanden werden können,

[3] https://computationalthinkingcourse.withgoogle.com/unit (Zugriff 13.04.2019)

sodass auch die Vermittlung von informatischen Methoden in Kombination mit Inhalten von anderen Fachbereichen erleichtert würde [Ar14]. Einige Probleme der Informatiklehre in Schulen, wie zum Beispiel die Wahrnehmung von Informatik als reine Informationstechnologie, könnten so überwunden werden. Einige Autor*innen haben dennoch die Umsetzung des Konzepts in den schulischen Curricula kritisiert [HMN17]. Insbesondere verbreite das Konzept unklare bzw. vieldeutige Definitionen und dogmatische Sichtweisen von Informatik, die andere Perspektiven exkludierten.

Mannila et al. bezeichnen das Konzept als unpräzise [Ma14]. Gudzial kritisiert die Idee „allgemeingültiger" Vorteile von Computational Thinking als übertrieben und weist darauf hin, dass es trotz der Popularität von Computational Thinking, noch keinen Konsens über die konkreten Inhalte gibt, die in den Schulen in den USA integriert werden sollen [Gu15]. Tedre und Denning verweisen in ihrer Kritik auf die unterschiedlichen Phasen der historischen Entwicklung der Informatik als Disziplin und als Profession in den USA [TD16]. Diese Autor*innen erkennen Wings Konzept als eine Wiederentdeckung von Paperts Ideen aus den achtziger Jahren des vergangenen Jahrhunderts über die Informatikausbildung in Schulen an.

Tedre und Denning identifizieren sechs konkrete Risiken, die mit der Umsetzung des Computational Thinking Konzepts in den Schulen verbunden sind: Mangel an Ehrgeiz, Dogmatismus, Betonung von Wissen statt Praxis, übertriebene Forderungen, beschränkte Sichtweisen über die Informatik und übertriebene Betonung von Formulierungen [TD16: 125 ff.]. Der visionäre Ehrgeiz während des letzten Jahrhunderts, die Informatikausbildung in Schulen einzuführen, wurde nach den Analysen dieser Autor*innen von Computational Thinking - Ideen ersetzt. Diese fokussieren sich lediglich auf Programmierungsmethoden, obwohl Digitalisierungsprozesse heutzutage erweiterte Sichtweisen jenseits von Methoden und Technologien erfordern.

Mit der Popularisierung von Computational Thinking gehe auch die Gefahr einher, dass eine dogmatische Sicht über die Informatik verankert würde, sodass alternative Sichtweisen exkludiert bzw. unberücksichtigt und unsichtbar blieben. Darüber hinaus sei bei der Definition von Computational Thinking der CSTA unklar, wie das anvisierte Wissen in Skills für die professionelle Praxis der Informatik übertragen werden könne. Welche konkreten Praktiken sollen also in das schulische Curriculum integriert werden, die zum Erlernen spezifischer Fähigkeiten führen?

In einigen skandinavischen Ländern wurden die Prinzipien des Computational Thinking unter der Rubrik „digitale Kompetenzen" konkreter und auch ehrgeiziger als in den USA in den schulischen Curricula integriert. In 2017 beispielsweise billigte die Schwedische Regierung die Integration von so genannten „digitalen Kompetenzen" in das Curriculum der Sekundärstufen, die ab Herbst 2018 unterrichtsverpflichtend wurde [HMN17] . Diese Kompetenzen basieren auf den, von der Europäischen Kommission formulierten, digitalen Kompetenzen[4]. Vier Prinzipien werden grundsätzlich für die Integration von digita-

[4] https://ec.europa.eu/jrc/en/digcomp (Zugriff 3.04.2019)

len Kompetenzen in den Sekundarstufen betont: 1) Verstehen, wie Digitalisierung Individuen und Gesellschaften beeinflusst, 2) Verstehen und Wissen darüber, wie digitale Technologien funktionieren, 3) kritische und verantwortungsvolle Nutzung von digitalen Technologien und Ressourcen, und 4) Fähigkeiten erlernen, um Probleme zu lösen und Ideen in der Praxis zu implementieren [HMN17][5]. Programmierung wird nicht nur als Codieren, sondern als ein Teilbereich digitaler Kompetenzen konzeptualisiert. Betont wird allerdings, dass Programmierung ein Prozess ist, der in verschiedene Phasen gegliedert ist und immer in einem breiten Kontext von Schöpfung, Kontrolle und Regulierung sowie Simulationen und demokratischen Dimensionen positioniert werden sollte. Basierend auf diesen vier Prinzipien werden digitale Kompetenzen nicht als ein separates Fach in schwedischen Schulen integriert, sondern als Teil verschiedener Schulfächer. Speziell in den sozialwissenschaftlichen Fächern werden konkrete Aspekte der digitalen Kompetenzen, wie bspw. eine verantwortungsvolle Nutzung von digitalen Medien aus einer sozialen, ethischen und legalen Perspektive, die Darstellung von Personen im Internet aus Genderperspektive, die Informationskontrolle durch versteckte Programmierung oder der Einfluss von Digitalisierungsprozessen auf unterschiedliche Aspekte von Gesellschaft integriert. In dem Konzept des Critical Computational Thinking, das ich im Folgenden skizziere, schlage ich vor, noch eine weitere Dimension zu integrieren: die Rolle von Informatikwissen in der Gestaltung von Digitalisierung. Das bedeutet einen Perspektivenwechsel über die Rolle der Informatik in Digitalisierungsprozessen – weg von der Rolle der neutralen Problemlöserin hin zur Reflexion der Beteiligung an der Gestaltung von Digitalisierungsprozessen durch die „Diagnose“ von Problemen sowie durch die Identifizierung von möglichen Problemerweiterungen bei der Umsetzung von „Therapielösungen“. Denn die Anwendung von Informatikmethoden bzw. –wissen setzt eine gewisse Realitätsselektion und Lösungspriorisierung voraus.

In diesem Zusammenhang schlage ich vor, Ethik- und Professionalitätsfragen der Informatik[6] bezogen auf Digitalisierungsprozesse (s. oben) in die Lehre zu integrieren, die in interdisziplinären Projekten nicht nur in Schulen sondern auch in weiteren Settings der Informatikdidaktik (z. B. in Universitäten und Fachhochschulen) umgesetzt werden können. In diesen Projekten können jenseits der Anwendung informatischer Methoden wie im Konzept des Computational Thinking vorgeschlagen wird, „Digitalisierungsprobleme“ definiert, modelliert, analysiert und evaluiert werden. In den nächsten Seiten wird dieses Konzept skizziert.

3 Critical Computational Thinking

In dem skizzierten Ansatz des Critical Computational Thinking nutze ich das Konzept von digitalen Kompetenzen als Erweiterung von Computational Thinking Prinzipien. Darüber hinaus schlage ich vor, dass eine ethische (kritische) Hinterfragung von Digita-

[5] https://link.springer.com/chapter/10.1007%2F978-3-319-71483-7_10 (Zugriff 5.04.2019)

[6] Speziell über die Fragen des Framing von Informatikbildung s. Schulte und Budde [SB18]

lisierung [Gr18] als integraler Teil der Informatikausbildung berücksichtigt wird. Dabei geht es nicht um die Technik an sich, sondern um digitale Techniken, Artefakte und Systeme in konkreten Kontexten. Folgende Prinzipen aus dem Computational Thinking Konzept bilden zunächst die Basis der Vermittlung von Informatikmethoden:

Computational Thinking Prinzipien
Muster erkennen (Beobachten und Ähnlichkeiten identifizieren)
Dekomponieren (In Teile Aufbrechen)
Abstrahieren (Auf unnötige Details verzichten)
Logisches Denken (Modellieren)
Algorithmen definieren (Teile aufeinander beziehen, Schritte und Regeln definieren)
Evaluieren (Urteilen). Simulationen

Tab 1: Prinzipien des Computational Thinking zur Vermittlung von Informatikmethoden

Informatikmethoden benötigen eine theoretische Begründung, eine Kontextualisierung sowie konkrete Fragen für ihre praktische Umsetzung. Das Konzept des Critical Computational Thinking umfasst ethische Prinzipien, die bereits in der Informatik diskutiert werden [Mi03], aber auch ethische Reflexionen, die durch die Entwicklung von Digitalisierungsprozessen entstanden sind [Gr18, Gru18]. In Anlehnung an Grunwald [Gru18: 5], der moralische Aspekte von Technik handlungstheoretisch benennt, können sich ethische Aspekte von Digitalisierungsprozessen auf die (mit digitalen Artefakten und Systemen) verfolgten Digitalisierungs**ziele**, auf die (zur Digitalisierung) eingesetzten **Mittel** sowie auf die **Folgen** (inkl. nicht intendierter Folgen) und Unsicherheiten von Digitalisierungsprozessen beziehen.

Digitalisierungsprozesse in unterschiedlichen Lebensbereichen (z. B. Arbeit, Gesundheit, Mobilität, etc.) bilden die Kontextualisierungsbeispiele für die Vermittlung von Informatikmethoden bzw. von den Prinzipien des Computational Thinking. Diese Digitalisierungsprozesse sind anhand Computational Thinking Prinzipien und mit konkreten, digitalen Kompetenzen gesellschaftlich, sozial und individuell gestaltbar. Das heißt, dass bereits in der Schule, Personen für Digitalisierungsgestaltung und für die Nutzung von digitalen Artefakten und Systemen befähigt werden könnten, sodass unreflektierte Faszination von Digitalisierung und Bequemlichkeiten bzw. Unmündigkeit in der Nutzung und in der Gestaltung von digitalen Artefakten und Systemen vermieden werden [Gru18].

Ethische Reflexionen über die Ziele, die eingesetzten Mittel sowie über die Folgen der Gestaltung und Nutzung von digitalen Techniken, Artefakten und Systemen in konkreten sozialen Kontexten (z. B. Arbeit, Gesundheit, Mobilität, etc.) sind im Konzept des Critical Computational Thinking Teil der digitalen Kompetenzen, die die Vermittlung von Informatikmethoden ergänzen. Damit wird erzielt, dass die Diskussion über informatische Kompetenzen und informatische Grundbildung in Schulen und weiteren Bereichen der Vermittlung von Informatikwissen im Zusammenhang mit Digitalisierungsprozessen, mit den unterschiedlich beteiligten Akteur*innen vorangebracht und kontextualisiert

wird (z. B. in Schulen: mit Lehrkräften aus verschiedenen Fachbereichen, Schuldirektor*innen, Eltern, Ministerien). Die existierenden Grundlagen, die von der Gesellschaft für Informatik (GI) sowie von der Kultusministerkonferenz (KMK) formuliert wurden [Ei18] definieren die „digitale vernetzte Welt" als etwas von außen Vorgegebenes, das lediglich zu nutzen ist und bloß in seinen Funktionen und Wirkungen zu verstehen sei. Damit wird vor allem die Rolle von Nutzer*innen in den Vordergrund gestellt. Digitale Medien sind dennoch nicht nur Hilfsmittel (zum Beispiel im Unterricht), sie bestimmen maßgeblich menschliche Wahrnehmungen und soziales Handeln mit [He16: 7]. Digitalisierungsprozesse sind grundsätzlich gestaltbare Prozesse, für die alle Teilnehmer*innen verantwortlich sind, auch wenn ihre Wirkungen nur teilweise vorhersehbar sind. Diese Gestaltung von Digitalisierungsprozessen geschieht nicht nur durch die Nutzung von existierenden digitalen Geräten, sondern auch durch die Planung, Produktion von materiellen digitalen Artefakten, durch die Modellierung von sozialen und gesellschaftlichen Beziehungen oder durch die infrastrukturellen Vernetzungen. Das Konzept des Critical Computational Thinking integriert deshalb digitale Kompetenzen nicht als Fähigkeiten, die Personen bzw. Schüler*innen die bloße Nutzung von digitalen Geräten ermöglichen sollen oder als Berücksichtigung von *„Wechselwirkungen der digitalen vernetzten Welt mit Individuen und der Gesellschaft" (s. GI Manifesto S. 3).* Critical Computational Thinking berücksichtigt ethische Fragen, die interdisziplinär aus unterschiedlichen Perspektiven in kontextualisierten Digitalisierungsprozessen zunächst geklärt werden sollen und mit Hilfe von Informatikmethoden konkret beantwortet werden können. Speziell für eine Informatikausbildung im Kontext von Digitalisierungsprozessen sind diese Fragen unverzichtbar, denn insbesondere professionelle Informatiker*innen sind sowohl Nutzer*innen als auch Gestalter*innen von diesen Artefakten mit ihren benötigten Infrastrukturen und Vernetzungen.

Digitalisierungsprozesse werden aber nicht nur durch die Informatik gestaltet. Alle gesellschaftlichen Bereiche werden in Deutschland von Digitalisierungsprozessen betroffen sein. Deswegen ist es wichtig, dass die oben genannten, ethischen Aspekte von Digitalisierungsprozessen (Ziele, Mittel, Folgen) mit ihren unterschiedlichen sozialen, kulturellen, ökologischen und ökonomischen Implikationen in verschiedenen Lebensbereichen interdisziplinär bereits in den Schulen thematisiert und in gemeinsamen Projekten mit Computational Thinking Prinzipien erfahrbar gemacht werden. Um gemeinsame Projekte in der Vermittlung von Informatikwissen zu gestalten empfiehlt es sich, Informatikmethoden und digitale Artefakte nicht in den Vordergrund zu stellen, sondern eher von den o. g. Fragen *„Was ist wünschenswert? und „Für wen ist das wünschenswert?"* [Gr18] in einem konkreten Kontext auszugehen (Beispiele wären Smart Homes, Smart Cities oder auch digitalisierte Pflege). Ungleiche Teilhabemöglichkeiten an Digitalisierungsprozessen sind mit ethischen Fragen insofern verbunden, als dass nicht alle Personen Gestaltungsfreiheit über den Einfluss der Digitalisierung in ihrem Leben haben. So zum Beispiel tragen Geschlechterkategorisierungen (in Interdependenz mit anderen sozialen Kategorisierungen wie z. B. Bildungsgrad oder die Einkommenshöhe) zur ungleichen Teilhabemöglichkeit an Digitalisierungsprozessen bei. Die Sichtbarmachung von ungleichen Teilhabemöglichkeiten, bezogen auf soziale Kategorisierungen wie z. B. Geschlecht bei den verschiedenen Phasen der Entwicklung digitaler Artefakte durch das

strategische Infrage stellen von Digitalisierungs**zielen**, -**mitteln** sowie **-folgen** (inkl. nicht intendierten Folgen) ist ein Beispiel für die kritische Reflexion, die im Konzept des Critical Computational Thinking integriert ist.

Konkret für die Vermittlung von Informatikwissen in Schulen sowie in anderen Settings der Informatikdidaktik im Kontext von Digitalisierungsprozessen, sind folgende Aspekte des skizzierten Critical Computational Thinking Konzeptes relevant:

Ethische Prinzipien Breites Verständnis von Digitalisierung, (Informatik) und digitalen Kompetenzen im Sinne des Gemeinwohls (*„Was ist wünschenswert? Für wen ist das wünschenswert?“* [Gr18])
Computational Thinking Prinzipien
Muster erkennen (Beobachten und Ähnlichkeiten identifizieren)
Dekomponieren (In Teilen Aufbrechen)
Abstrahieren (Auf unnötige Details verzichten)
Logisches Denken (Modellieren)
Algorithmen definieren (Teile aufeinander beziehen, Schritte und Regeln definieren)
Evaluieren (Urteilen). Simulationen
Digitale Kompetenzen
Umsetzung einer interdisziplinären kritischen Ethik-Reflexion der Gestaltung, Nutzung und Auswirkungen digitaler Artefakte.
(Diagnose, Inferenz, Behandlung). Welche „Probleme“ sind in Digitalisierungsprozessen relevant und warum? (Diagnose) Welche Informatikmethoden bzw. digitalen Tools können umgesetzt werden, um diese Probleme zu lösen? (Inferenz, Behandlung) Welche möglichen zusätzlichen „Probleme“ treten bei der Anwendung dieser Methoden auf?

Tab. 2: Skizzierte Aspekte von Critical Computational Thinking für die Informatikdidaktik im Kontext von Digitalisierungsprozessen

Critical Computational Thinking könnte zum Beispiel in interdisziplinären Projekten über Digitalisierung und Nachhaltigkeit umgesetzt werden. Erfahrungen, die bereits Easterbrook [Ea14] aus seinem Projekt gemacht hat, könnten zunächst ergänzt und an den konkreten Bildungskontext angepasst werden. Im Rahmen des BMBF Projektes Fix-IT werden Aspekte von Critical Computational Thinking in Workshops (z. B. über Hacken als Beruf) umgesetzt und evaluiert. In diesem Artikel kann jedoch noch nicht über die Umsetzung dieses Konzeptes berichtet werden.

4 Konklusion

In diesem Beitrag habe ich ein neues Konzept zur kritischen, ethisch-reflexiven Vermittlung von Informatikkenntnissen für die Gestaltung von Digitalisierungsprozessen skizziert, das anlehnend an Aspekte der Technikethik [Gr18, Gru 13], Prinzipen von Computational Thinking und digitale Kompetenzen im interdisziplinären Dialog mit verschiedenen Schulfächern kombiniert. In diesem Konzept des Critical Computational Thinking werden ethische Aspekte der Gestaltung und Nutzung von digitalen Artefakten betont und die Vermittlung von Informatikwissen als interdisziplinäres Projekt im Kontext von Digitalisierungsprozessen verstanden.

Die kritische reflexive Verbindung von ethischen Fragen mit der Umsetzung von Prinzipien des Computational Thinking und digitalen Kompetenzen können die Grundlage sein für die Entwicklung eines neuen Selbstverständnisses als informatische Lehrkraft und Angehörige*r einer Fachkultur sowie einer Öffnung eben dieser Fachkultur für Perspektiven, die bisher nicht oder nur marginalisiert in das Feld „der Informatik“ eingebunden waren. Durch die interdisziplinäre und kontextualisierte Vermittlung von Informatikwissen in Digitalisierungsprozessen, kann die Informatik sich selbst als Fachbereich neu definieren. So können die Informatiklernenden und werdende IT Expert*innen, unabhängig von ihrer sozialen und ökonomischen Verortung, auf die Herausforderungen von Digitalisierungsprozessen vorbereitet werden, um die Potentiale digitalisierter Arbeits- und Lebenswelten aktiv mitgestalten zu können.

Literaturverzeichnis

[Ab88] Abbott, A. 1988, The system of professions : an essay on the division of expert labor. Chicago: University of Chicago Press.

[Ar14] Arraki, K. 2014, DISSECT: An experiment in infusing computational thinking in K-12 science curricula. IEEE Frontiers in Education Conference (FIE) Proceedings.

[Auf97] Aufenanger, S., 1997, Medienpädagogik und Medienkompetenz. Eine Bestandsaufnahme. In: Enquete-Kommission „Zukunft der Medien in Wirtschaft und Gesellschaft. Deutschlands Weg in die Informationsgesellschaft“. Deutscher Bundestag (Hrsg.) Medienkompetenz im Informationszeitalter, Bonn, S. 15-22.

[BHC11] Barr, D., J. Harrison, and L. Conery, 2011. Computational thinking: A digital age skill for everyone. *Learning & Leading with Technology.* 38, no. 6: 20-23.

[De10] Dewey, J. 2010, *How we think*. Boston: D.C. Heath.

[Ea14] Easterbrook, S. 2014, From Computational Thinking to Systems Thinking. Proceedings of the 2nd International Conference on Information and Communication Technologies for Sustainability (ICT 4S2014).

[Ei18] Eickelmann, B. 2018, Digitalisierung in der schulischen Bildung. Entwicklungen, Befunde und Perspektiven für die Schulentwicklung und die Bildungsforschung. In: McElvany, N. et al. (Hrsg.) Digitalisierung in der schulischen Bildung: Chancen und

Herausforderungen. Waxmann.

[Fa] Facione, P. A. 1990, Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Asessment and Instruction. ERIC Document ED315423

[GI08] Gesellschaft für Informatik (Arbeitskreis »Bildungsstandards«) (2008) Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sekundarstufe I. https://www.informatikstandards.de/docs/bildungsstandards_2008.pdf

[Go11] Computational Thinking Course with Google https://computationalthinkingcourse.withgoogle.com/unit (Zugriff 13.04.2019)

[Gr18] Grimm, P. 2018, Grundlagen für eine digitale Wertekultur. In: Wolfgang Stadler (Hrsg.): Mehr als Algorithmen. Digitalisierung in Gesellschaft und Sozialer Arbeit. Sonderband TUP - Theorie und Praxis https://www.awo.org/sites/default/files/2018-09/TUP-Sonderband_2018_Grimm.pdf

[Gru18] Grunwald, A. 2018, Der unterlegene Mensch: Die Zukunft der Menschheit im Angesicht von Algorithmen, künstlicher Intelligenz und Robotern. Riva.

[Gru13] Grunwald, A. 2013, Handbuch Technikethik. Berlin: Springer.

[Gu15] Guzdial, M. 2015, Software realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments*, 4(1), 001–044.

[He16] Hessen, J. 2016, Handbuch Medien- und Informationsethik. Berlin: Springer.

[KB13] Kafai, Y. B., Burke, Q. 2013, The social turn in K-12 programming: moving from computational thinking to computational participation, Proceeding of the 44th ACM technical symposium on Computer science education, March 06-09, 2013, Denver, Colorado, USA [doi>10.1145/2445196.2445373]

[HMN17] Heintz, F., Mannila, L., Nordén, L.-A., Parnes, P., Regnell, B. 2017, Introducing Programming and Digital Competences in Swedish K-9 Education. In: *International Conference of Informatics in Schools: Situation, Evolution and Perspectives.ISSEP 2017: Informatics in Schools: Focus on Learning Programming. Pp.: 117-128.*

[Ma14] Mannila, L. 2014, Computational Thinking in K-9 Education. In: *ITiCSE-WGR*'14,June 23-25, 2014, Uppsala, Sweden

[Mi03] Mieth, D. 2003, Ethik der Informatik. In: INFORMATIK 2003 - Innovative Informatikanwendungen, Band 2, Beiträge der 33. Jahrestagung der Gesellschaft für Informatik e.V. (GI), 29. September - 2. Oktober 2003 in Frankfurt am Main.

[Pa80] Papert, S. 1980, *Mindstorms: Children, Computers, and Powerful Ideas.* Basic Books.

[Ru05] Ruiz Ben, E. 2005, Professionalisierung der Informatik. Wiesbaden: DUV Verlag.

[Ru18] Ruiz Ben, E. 2018, Intersectionality in the Practice of Mix-Methods Gender Research. In: *Journal of Research in Gender Studies. 8(1): 73–88.*

[SB18] Schulte, C., Budde, L. 2018, A Framework for Computing Education: Hybrid Interaction system. 18th Koli Calling Conference on Computing Education Research , No-

vember 22-25, 2018, Koli, Finland.

[TD16] Tedre, M. and Denning, P. J. 2016, The Long Quest for Computational Thinking. Proceedings of the 16th Koli Calling Conference on Computing Education Research , November 24-27, 2016, Koli, Finland: pp. 120-129.

[Wi06] Wing, J. M. 2006. Viewpoint - Computational thinking. *Communications of the ACM*. 49, no. 3: 33.

Informatik für alle – Eine Analyse von Argumenten und Argumentationsschemata für das Schulfach Informatik

Stefan Seegerer[1] Tilman Michaeli[2] Ralf Romeike[3]

Abstract: Die Wissenschaft Informatik ist die maßgebliche Triebkraft der sogenannten digitalen Transformation. Nichtsdestotrotz wird die Debatte um digitale Bildung in der Schule oftmals von „digitalen Medien" dominiert, und Informatik ist nur in wenigen Bundesländern ein Pflichtfach. Es ergibt sich damit immer wieder die Notwendigkeit, den Beitrag eines Schulfachs Informatik herauszustellen. Es gibt eine Vielzahl an Positionspapieren, wissenschaftlichen Beiträgen, Zeitungsartikeln, usw., die für eine verpflichtende curriculare Verankerung des Fachs argumentieren. In diesem Beitrag werden auf einer Datengrundlage von 50 solcher Dokumente aus dem deutschsprachigen Raum prävalente Argumente und deren Argumentationsschemata mit Hilfe einer qualitativen Inhaltsanalyse nach Mayring untersucht. Hierzu wird eine bestehende Einteilungen der Argumente und ein Kategoriensystem für Argumentationsschemata aus der Argumentationstheorie eingesetzt. Dabei zeigt sich, dass sich die Argumentation auf Verständnis und Teilhabe an der „digitalen Welt" konzentriert. Betrachtet man die Argumentationsschemata, ist festzustellen, dass insgesamt nur selten evidenzbasiert argumentiert wird, sondern vorwiegend auf Basis von Analogien oder allgemein akzeptierter Aussagen.

Keywords: Informatik für alle Schulfach Pflichtfach Argumente Argumentationsschemata Begründungen

1 Einleitung

Die Digitalisierung und die daraus resultierende digitale Transformation hat, obgleich nicht neu, in den letzten Jahren starken Einfluss auf formale Bildungsprozesse (vgl. etwa [St16]). Die zugrunde liegende Wissenschaft und zentrale Triebkraft dieser Entwicklungen ist die Informatik und Informatiksysteme sind in ihrer gesellschaftlichen und auch wirtschaftlichen Bedeutung unstrittig.

Während die Dynamik der Digitalisierung immer weiter zunimmt, ist ein verpflichtender Informatikunterricht *für alle* noch immer in weiter Ferne. Die Debatte um digitale Bildung in der Schule wird von Medienbildung und dem Umgang mit digitalen Medien dominiert (vgl. [Br17], [SS15]). In Zeiten, in denen informatische Bildung auch vermehrt in Grundschulkontexten diskutiert wird (etwa [Be17]), muss die Informatik immer noch

[1] Friedrich-Alexander-Universität Erlangen-Nürnberg, Didaktik der Informatik, Martensstr. 3, 91058 Erlangen, Deutschland, stefan.seegerer@fau.de
[2] Friedrich-Alexander-Universität Erlangen-Nürnberg, Didaktik der Informatik, Martensstr. 3, 91058 Erlangen, Deutschland, tilman.michaeli@fau.de
[3] FU-Berlin, Didaktik der Informatik, Königin-Luise-Str. 24-26, 14195 Berlin, Deutschland, ralf.romeike@fu-berlin.de

doi:10.18420/inf2019_77

ihren Stellenwert als eigenständiges Schulfach rechtfertigen. Die Diskussion begleitet die Informatik seit ihren Anfängen als Unterrichtsfach (vgl. [Ke90]) bis heute [BM19] und ist nicht auf bestimmte Regionen beschränkt (etwa [Je10], [Be03] oder [Fl16]). Welche Argumente werden in dieser Diskussion angeführt? Und auf welche Art und Weise wird argumentiert? Bisher existieren lediglich wenige empirische Untersuchungen über die in der Diskussion um Informatik als Schulfach verwendete Argumentationsführung.

Dieser Beitrag untersucht die wesentlichen Argumente, die in der Rechtfertigung und Diskussion im deutschsprachigen Raum Verwendung finden und Informatik als Pflichtfach rechtfertigen sollen. Darüber hinaus wird auch die Art und Weise der Argumentation analysiert.

2 Hintergrund

Bildung bedeutet die „Befähigung zur aktiven Beteiligung am beruflichen und öffentlichen Leben“[GF09]. (Allgemein-)Bildung kann dabei verschiedene Dimensionen haben, so schlüsselt beispielsweise die Gesellschaft für Fachdidaktik den Bildungsbegriff in die Bereiche Identitätsbildung, Alltagsbewältigung, Ausbildungsreife und Partizipation auf [GF09]. Heymann entwickelte daher mit den sieben Aufgaben allgemeinbildender Schulen einen Maßstab, der herangezogen werden kann, um den allgemeinbildenden Wert eines Faches herauszustellen [He96]. Witten hat mithilfe dieses Katalogs den Informatikunterricht untersucht [Wi03].

Betrachtet man die Debatte um das Schulfach Informatik, zeigt sich, dass die Diskussion dort auch Argumente mit einbezieht, die nicht in direkter Verbindung mit Heymanns Katalog stehen, etwa im Bezug auf den Arbeitsmarkt. Es gab und gibt immer wieder Ansätze die Argumentation um Informatik als Schulfach zu strukturieren.

Döbeli Honegger fasst die Argumentation für Informatik in der Schule und damit deren Allgemeinbildungsanspruch in neun Argumenten zusammen: Konstruktionismusargument, Wissenschaftsargument, Denkobjektargument, Problemlöseargument, Arbeitstechnikargument, Interesseargument, Berufswahlargument, Welterklärungs- oder Mündigkeitsargument und Konzeptwissenargument [DH16].

Passey [Pa17] identifiziert sechs Argumente: Informatische Bildung für alle soll dafür sorgen, dass zukünftige Anforderungen der Wirtschaft erfüllt (*economic argument*), (soziale) Interessensgemeinschaften unterstützt (*community argument*), Verständnis geschaffen (*educational argument*), Fähigkeiten wie Problemlösen ausgebildet (*learning argument*) und Teamwork geschult wird (*organisational argument*). Zudem führt er an, dass Lernende auch die Möglichkeit haben sollten, sich mit den Dingen zu beschäftigen, die sie selbst interessieren (*learner argument*).

Vogel et al. hingegen wählen einen empirischen Ansatz: In einer dreistündigen Arbeitsgruppen forderten sie 24 Experten auf, Argumente für informatische Bildung niederzuschreiben.

In den 161 genannten Einzelargumenten und projizierten Auswirkungen informatischer Bildung identifizierten sie sieben Bereiche, die „CSed Visions“ [VSC17]: *economic and workforce development*, *equity and social justice*, *competencies and literacies*, *citizenship and civic life*, *scientific, technological and social innovation*, *school improvement and reform* sowie *fun, fulfillment and personal agency*.

Eine Argumentation erfolgt dabei auf zwei Ebenen: einer inhaltlichen Ebene (Argument) und der Art und Weise, wie die Argumentation geführt wird (Argumentationsschema).

- Ein *Argument* besteht immer aus mehreren Teilen: Der Konklusion, die die zu begründende Behauptung darstellt und einer beliebigen Anzahl von Prämissen, die diese Behauptung begründen sollen [Ba07].
- *Argumentationsschemata* beschreiben nun die Struktur der Prämissen, also auf welche Art und Weise die Konklusion begründet wird. So kann eine Konklusion etwa durch eine Analogie, eine allgemein akzeptierte Aussage oder die Meinung eines Experten begründet werden.

3 Forschungsfragen

Ziel dieser Untersuchung ist es, die Argumente und Argumentationsschemata der Diskussion (im deutschsprachigen Raum) in Dokumenten wie Positionspapieren, Essays oder wissenschaftlichen Beiträgen um ein verpflichtendes Schulfach Informatik zu analysieren. Damit werden die folgenden Forschungsfragen adressiert:

- (RQ0) Welche Argumente werden für ein Schulfach Informatik angeführt?
- (RQ1) Welche Argumentationsschemata werden verwendet?

4 Methodik

Das Vorgehen zur Beantwortung dieser Forschungsfragen orientiert sich an den Schritten der strukturierenden qualitativen Inhaltsanalyse nach Mayring [Ma00]. Die qualitative Inhaltsanalyse beschreibt ein Verfahren zur systematischen Textanalyse. Ziel einer solchen ist die Abbildung der wesentlichen Aspekte des untersuchten Materials.

Materialauswahl. Für die Analyse wurden Dokumente aus dem deutschsprachigen Raum herangezogen, die für Informatik in der Schule argumentieren. Dazu wurden die Seiten der GI sowie ihrer Landesgruppen und einschlägige Onlinebibliotheken (Springer Link, Google

	Kategorie	Beschreibung
Reasoning	*Deductive Reasoning*	Aus einer oder mehreren Aussagen (Prämissen) wird eine logisch sichere Schlussfolgerung abgeleitet.
	Inductive Reasoning	Eine oder mehrere Aussagen werden als Beweis für die Wahrheit der Schlussfolgerung angesehen (Gegenbegriff zu *Deductive Reasoning*).
	Practical Reasoning	Ausgehend von einem Ziel wird argumentiert, dass eine Aktion erforderlich ist, um dieses zu erreichen. Dazu gehören das Ausschließen von Alternativen (*Argument from Alternatives*), das Skizzieren von Gefahren (*Argument from Threat*) oder Nennen der folgenden Konsequenzen (*Argument from Consequences*).
	Abductive Reasoning	Ausgehend von einer überraschenden Beobachtung wird eine Erklärung gebildet.
	Causal Reasoning	Ausgehend von einer Ursache und ihrer Wirkung wird eine Kausalität identifiziert.
Source-based Arguments	*Argument from Position to Know*	Das Argument stützt sich auf eine Quelle, die über entsprechendes Wissen verfügt (etwa eine Expertin (*Argument from Expert Opinion*)), oder nutzt Aussagen, die als wahr zu akzeptieren sind, solange nicht das Gegenteil bewiesen wird (*Argument from Ignorance*).
	Argument from Commitment	Argument, bei dem sich auf eine entsprechend bestätigende Aussage des Angesprochenen bezogen wird.
	Arguments Attacking Personal Credibility	Argument, bei dem die Position eines anderen Diskussionsteilnehmers aufgrund persönlicher Eigenschaften angefochten wird (z.B. Voreingenommenheit).
	Arguments from Popular Acceptance	Argument, dass eine Aussage allgemein akzeptiert wird und dass sie daher (vorläufig) als plausibel akzeptiert werden kann [Wa05a].
Applying Rules to Cases	*Defeasible Rule-Based Arguments*	Die Argumentation nutzt regelbasierte Schemata, ist aber anfechtbar.
	Arguments Based on Cases	Das Argument stützt sich auf Beispiele (*Argument from Example*), Analogien (*Argument from Analogy*) oder Präzedenzfälle (*Argument from Precendent*).
	Verbal Classification Arguments	Pauschalurteil, bei dem etwas pauschal in eine Kategorie eingeordnet wird [An16].
	Chained Arguments Connecting Rules and Cases	Argumentationsweise, bei der eine Reihe von aufeinander aufbauenden Konsequenzen angeführt wird.

Tab. 1: Übersicht über Argumentationsschemata

Scholar und die digitale Bibliothek der GI) auf entsprechende Dokumente bzw. nach Stichwörtern wie „Informatik für alle“, „Pflichtfach Informatik“ oder „Informatische Bildung“ untersucht. Ergänzt wurde dies durch eine Webrecherche nach denselben Schlagworten. Wie sich die resultierenden 50 Dokumente nach Jahr der Veröffentlichung aufteilen findet sich in Abbildung 1.

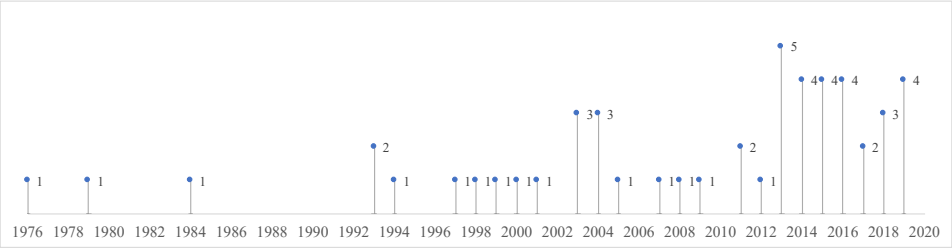


Abb. 1: Anzahl an Quellen nach Jahr der Veröffentlichung

Kategoriensystem. Mayring unterscheidet zwischen zwei Verfahren zur Bildung eines Kategoriensystems: Zunächst kann ein Kategoriensystem induktiv auf Basis der Datengrundlage gebildet oder aber deduktiv, z.B. aus der Literatur, abgeleitet werden. Die zu kodierenden Dokumente werden dann im qualitativen Analyseschritt den entsprechenden Kategorien zugeordnet.

Für die Analyse der Argumentation auf *inhaltlicher Ebene* (RQ0) haben wir ausgehend von den Argumentationssammlungen von Döbeli Honegger [DH16], Passey [Pa17] und Vogel et al. [VSC17] deduktiv ein Kategoriensystem entwickelt, indem deren Kategoriensysteme vereinigt wurden. Die Argumentationen von Passey, Döbeli Honegger und Vogel et al. und das resultierende Kategoriensystem werden in Tabelle 2 gegenübergestellt. Um wichtige Aspekte nicht durch vordefinierte Kategorien zu vernachlässigen, waren induktive Ergänzungen von Kategorien zulässig.

Für die Analyse der *Argumentationsschemata* (RQ1) haben wir das Modell von Walton, Reed und Macagno [WRM08] aus der Argumentationsforschung, ein häufig genutztes Instrument zur Klassifikation von Argumentationsschemata, als Kategoriensystem herangezogen. Argumentationsschemata werden dort zunächst in Schlussfolgern (*Reasoning*), quellbasierte Argumente (*Source-based Arguments*) und Anwenden von Regeln (*Applying Rules to Cases*) unterteilt (für die Unterkategorien siehe Tabelle 1).

Für die Zuordnung von Textstellen zu Kategorien wurde ein Kodierleitfaden erstellt (exemplarisch dargestellt in Tabelle 3).

Kodierungsphase. Nach Auswahl des Materials und Festlegung des Kategoriensystems bzw. der Kodierungskriterien erfolgte die eigentliche Analyse mithilfe der Analysesoftware MaxQDA. Im Idealfall konnte jede neu auftretende Kodiereinheit einer der bereits bestehen-

	Vogel et al. [VSC17]	Döbeli Honegger [DH16]	Passey [Pa17]
Berufs- und Arbeitswelt	economic and workforce development	Berufswahlargument	economic argument
Chancengleichheit	equity and social justice		community argument
Problemlösen	competencies and literacies	Problemlöseargument	learning argument
Überfachliche Kompetenzen		Arbeitstechnikargument	organisational argument
Konzeptwissen		Konzeptwissenargument	
Verständnis und Teilhabe	citizenship and civic life	Denkobjektargument	
		Welterklärungs- bzw. Mündigkeitsargument	educational argument
Wissenschaft	scientific, technological and social innovation	Wissenschaftsargument	
Positiver Einfluss auf Schule	school improvement and reform	Konstruktionismusargument	
Interessant und erfüllend	fun, fulfillment and personal agency	Interesseargument	learner argument

Tab. 2: Übersicht über verschiedene Argumente

den Kategorien zugeordnet werden. War dies nicht der Fall, so wurde – zumindest temporär – eine eigene Kategorie angelegt. Zum Abschluss der Analyse eines Dokuments wurden neu entstandene Kategorien erneut betrachtet. Konnte erneut keine Zuteilung zu einer der bestehenden Kategorien gefunden werden, so wurde diese als zusätzliche Kategorie aufgenommen.

Definition	Prototypische Textstelle	Kodierkriterium
Verständnis und Teilhabe	Die Kenntnis, Anwendung und kritische Reflexion der grundlegenden Konstruktionsprinzipien von Informationssystemen dient daher der Lebensvorbereitung und der Orientierung in einer von diesen Systemen geprägten Welt.	Die Konklusion des Arguments besagt, dass sich die digitale Welt/Informatiksysteme/... nur mit Informatik verstehen und mitgestalten lassen, ein mündiger Bürger Informatik benötigt oder Informatik hilft, grundlegende Vorstellungen (etwa von Intelligenz) auszuschärfen.
Argument from Popular Opinion	Unsere Gesellschaft befindet sich an der Schwelle des Übergangs von der Industriegesellschaft zur Informations- und Wissensgesellschaft ...	Eine allgemein akzeptierte Aussage wird zur Unterstützung des Arguments verwendet. [Wa05b, S. 91]

Tab. 3: Auszug aus dem Kodierleitfaden

5 Ergebnisse

Im Folgenden werden die einzelnen identifizierten Argumente (RQ0) und die jeweils zugehörigen Argumentationsschemata (RQ1) ausgeführt.

5.1 Berufs- und Arbeitswelt

Argument. 33 von 50 Dokumenten stützen sich auf das Berufs- und Arbeitsweltargument, das Anforderungen des Arbeitsmarktes in den Mittelpunkt stellt. Es finden sich dabei mehrere Aspekte: Zum einen wird argumentiert, Informatik helfe den Fachkräftemangel zu lindern, etwa indem mehr Personen für ein Studium in diesem Bereich gewonnen werden. Zugleich wird argumentiert, sie sei notwendig, um den Wirtschaftsstandort zu erhalten bzw. die Wettbewerbsfähigkeit auch in Zukunft sicherzustellen. Ein realistisches Bild helfe zudem Studienabbrüche zu verhindern. Eine weitere Facette des Arguments beleuchtet eine persönliche Perspektive: Zukünftig würde jeder informatische Kompetenzen in allen Berufen benötigen. Damit bereite informatische Bildung auch auf die eigene Berufsausbildung bzw. das Studium vor.

Argumentationsschemata. Die Argumentation beim Berufs- und Arbeitsweltargument ist stark praktisch geprägt: Vielfach wird ausgehend von einer Drohsituation argumentiert, die es zu vermeiden gelte (*Argument from Threat*). Es gelte, dem Arbeitsmarkt genügend qualifizierte Arbeitskräfte zur Verfügung zu stellen oder Studienabbrüche zu vermeiden. Auch die weitere Argumentation bedient sich praktischer Gründe: Der wirtschaftliche Gesamtschaden falle größer aus, als die jetzt notwendigen Investitionen in die Lehrkräftebildung (*Argument from Consequences*). Gleichzeitig wird auch mithilfe allgemein akzeptierter Aussagen insbesondere hinsichtlich des Fachkräftemangels argumentiert (*Argument from Popular Acceptance*). Zumindest der Werbeeffekt für ein Informatikstudium oder der Fachkräftemangel werden zudem durch Studien belegt (*Inductive Reasoning*).

5.2 Chancengleichheit

Argument. Das Argument der Chancengleichheit betont, dass informatische Kompetenzen – als wichtige Voraussetzungen für Teilhabe in der „digitalen Welt“ – unabhängig von Geschlecht, Herkunft oder sozialen Umständen zugänglich sein müssen. Es wurde in 18 von 50 Dokumenten kodiert, dabei werden in der deutschsprachigen Diskussion insbesondere zwei Aspekte hervorgehoben. Zunächst wird betont, wie wichtig verpflichtende informatische Bildung sei, um bereits früh Interesse zu wecken und so insbesondere Mädchen vor der Pubertät zu erreichen. Zum Anderen wird argumentiert, dass freiwillge Angebote nur Eliten erreichen und Maßnahmen nötig seien, sodass *alle* davon profitieren können.

Argumentationsschemata. Für das Argument der Chancengleichheit wird aus der praktischen Notwendigkeit heraus argumentiert, dass gewisse Zielgruppen nicht erreicht werden (*Argument from Consequences*). So würde eine „Digitale Spaltung“ drohen (*Argument from Threat*). Gleichzeitig werden Statistiken zum Frauenanteil in informatiknahen Berufen zum Unterstreichen der Argumentation herangezogen (*Inductive Reasoning*).

5.3 Denkweisen

Argument. Das Argument stellt genuin informatische Denkweisen heraus, die Lernende zur effektiven Lösung von Problemen innerhalb der Informatik und darüber hinaus befähigen und wird in 27 Dokumenten ausgeführt. Auch dieses Argument hat mehrere Facetten: Zum einen wird hervorgehoben, dass Informatik sich durch die Kombination mathematischen und ingenieurwissenschaftlichen Problemlösens von anderen Denkweisen abhebt. Zum anderen wird betont, dass damit eine universelle Problemlösekompetenz gefördert wird, die dann in anderen Bereichen angewendet werden kann. Auch der Begriff des Computational Thinking wird hier angeführt.

Argumentationsschemata. Mit informatischen Denkweisen wird ausgehend von Alternativen argumentiert (*Argument from Alternatives*): Nur dedizierte informatische Bildung vermittle diese Denkweisen, andere Fächer seien dazu nicht in der Lage. Gleichzeitig wird auch mit Hilfe von Analogien argumentiert: Auch die Mathematik habe anerkannte grundlegende Denkweisen, dasselbe gelte auch für die Informatik (*Argument from Analogy*). Teilweise wird das Argument zusätzlich, teilweise auch ausschließlich mit Beispielen gestützt, etwa welche Problemlösestrategien die Informatik nutzt und wo sich Computational Thinking in anderen Fächern zeigt (*Argument from Example*).

5.4 Überfachliche Kompetenzen

Argument. Das Argument besagt, dass die Informatik mit ihren Methoden und Arbeitsweisen zum Erwerb überfachlicher Kompetenzen beitrage. 17 von 50 Dokumenten stützen sich darauf. Dabei werden im Wesentlichen zwei Aspekte ausgeführt: So würde Informatik zur Förderung von Sozialkompetenz und Selbstorganisation beitragen. Seltener wird angeführt, dass informatische Bildung Kreativität fördere.

Argumentationsschemata. Die Förderung von Sozialkompetenz wird durch Projekte als aus der Praxis und Wissenschaft stammende und typische Methode im Informatikunterricht begründet. Weitere Begründungen, insbesondere Studien, inwiefern informatische Bildung überfachliche Kompetenzen unterstützt, werden nicht angeführt. Die Argumentation nutzt zu einem großen Teil allgemein akzeptierte Aussagen zu Projektarbeit (*Argument from Popular Acceptance*).

5.5 Konzeptwissen

Argument. 22 Dokumente argumentieren, dass informatische Bildung für die kompetente Nutzung digitaler Medien bzw. Werkzeuge unerlässlich sei. Diese böte langfristig anwendbares Konzeptwissen, welches nicht an konkrete Produkte oder Medien gekoppelt sei.

Argumentationsschemata. Die Argumentation baut meist auf allgemein akzeptierten Aussagen über Digitalisierung auf (*Argument from Popular Acceptance*). Gleichzeitig werden Alternativen, etwa ein Computerführerschein oder eine ähnliche Anwenderschulung, als unzureichend beschrieben (*Argument from Alternatives*). Zusätzlich werden Beispiele skizziert, für die informatische Kenntnisse benötigt werden, etwa beim Ergreifen von Sicherheitsvorkehrungen in Rechnernetzen (*Argument from Example*).

5.6 Verständnis und Teilhabe

Argument. Das Argument betont die Bedeutung informatischer Bildung für Verständnis und Teilhabe an der „digitalen Welt“. Dieses Argument wird in 48 von 50 Dokumenten verwendet. Es zeigt sich in verschiedenen Ausprägungen. In seiner einfachen Form beschränkt sich das Argument darauf, dass informatische Bildung notwendig ist, um die digitale Welt zu verstehen. Erst dann sei es möglich, mündig zu handeln und den digitalen Wandel selbst zu gestalten. Das fast nie genutzte Denkobjektargument betont die Möglichkeit, mit diesem Wissen über Informatik das eigene Verständnis über (menschliche) Konzepte wie Intelligenz zu schärfen.

Argumentationsschemata. Im Kontext des Verständnis- und Teilhabe-Arguments wird vorwiegend über Analogien argumentiert: Genau wie die Chemie oder die Physik bestimmte Aspekte der Welt erkläre, entmystifiziere auch die Informatik einen Teil (*Argument from Analogy*). Seltener erfolgt eine fallbasierte Argumentation über Beispiele (*Argument from Example*). Es wird argumentiert, dass die bestehenden Alternativen keinen ähnlichen Beitrag zur Welterklärung leisten können (*Argument from Alternatives*). Zudem werden Experten zitiert (*Argument from Expert Opinion*) oder mögliche Konsequenzen ausbleibender informatischer Bildung skizziert (*Argument from Consequences*).

5.7 Wissenschaft

Argument. Das Wissenschaftsargument fußt auf der Bedeutung der Informatik als Innovationstreiber in vielen anderen Bereichen und besagt, dass sich mit informatischen Methoden, insbesondere Simulationen oder Datenanalysen, neue wissenschaftliche Erkenntnisse gewinnen lassen. Lediglich 8 Dokumente verwenden dieses Argument.

Argumentationsschemata. Die Argumentation stützt sich auf allgemein akzeptierte Aussagen über die stetig steigende Datenmenge (*Argument from Popular Acceptance*). Gleichzeitig werden oft Beispiele angeführt, in denen informatische Methoden zu einem Erkentnissgewinn in den unterschiedlichsten Wissenschaften beitragen (*Argument from Example*). Dabei gebe es keine Alternative, denn das Potenzial könne ausschließlich mit ausreichend informatischer Bildung genutzt werden (*Argument from Alternatives*).

5.8 Positiver Einfluss auf Schule und Lernen in anderen Fächern

Argument. 30 von 50 Dokumenten begründen die Bedeutung informatischer Bildung in ihrem positiven Einfluss auf das Lernen in anderen Fächern bzw. die Institution Schule im

Allgemeinen. In seiner häufigsten Ausprägung wird erläutert, dass Informatik ein wichtiges systematisierendes und vernetzendes Element schulischer Bildung darstelle. Darüber hinaus wird betont, dass Informatik das Lernen in anderen Fächern unterstützt bzw. Gelerntes in anderen Fächern angewendet werden kann. Gleichzeitig sei Informatik notwendig, um Informatiksysteme in anderen Fächern adäquat nutzen zu können. Weiterhin wird der positiven Einfluss von Informatik auf Schule im Allgemeinen beschrieben, da nur Informatik bestimmte, wichtige Aufgaben schulischer Bildung erfüllen könne.

Argumentationsschemata. In Verknüpfung mit dem Konzeptwissenargument wird argumentiert, dass nur die Informatik eine entsprechende Handlungskompetenz im Umgang mit Informatiksystemen in anderen Fächern schaffen könne (*Chained Argument*). Außerdem gäbe es keine Alternativen, da nur Informatik digitale Medien zum Gegenstand des Unterrichts mache (*Argument from Alternatives*). Um zu begründen, dass Informatik ein systematisierendes Element schulischer Bildung darstellt, wird die Analogie zur Mathematik gesucht (*Argument from Analogy*) oder dies anhand von Beispielen expliziert (*Argument from Example*).

5.9 Interessant und erfüllend

Argument. Das Argument stellt den Lernenden in den Mittelpunkt: So wird dargelegt, dass die Lernenden selbst Interesse an der Informatik zeigen und wissen wollen, wie etwas funktioniert. Weiterhin sei die Informatik eine attraktive Disziplin: Es mache Spaß und sorge für persönliche Zufriedenheit und bereichere damit die Ausbildung. Dieses Argument findet sich in 12 untersuchten Dokumenten.

Argumentationsschemata. Die Argumentation beruht hauptsächlich auf allgemein akzeptierten Aussagen oder Meinungen (*Argument from Popular Acceptance*). In einigen Fällen wird auch ausgehend von Beispielen, etwa aus der Programmierung, argumentiert, aus denen die Faszination der Lernenden oder die Möglichkeiten zur persönlichen Entfaltung hervorgehen sollen (*Argument from Example*). In einigen wenigen Fällen wird das Argument mit konkreten Zahlen zur Belegung von Wahlfachunterricht im Bereich Informatik unterstrichen (*Inductive Reasoning*).

5.10 Weitere Argumente

Obwohl mit den ursprünglichen Kategorien bereits ein großer Teil der vorgebrachten Argumente erfasst werden konnte, ergab sich während der Analyse die Notwendigkeit zusätzliche Kategorien einzuführen. Das „Andere Länder machen es vor“-Argument beschreibt, dass informatische Bildung für alle in anderen Ländern bereits Normalität ist

(*Arguments Based on Cases*). Das „Die Zeit ist reif"-Argument betont, dass Informatik auf gesicherten Erkenntnissen aufbauen kann und es mittlerweile genügend Vorarbeiten für die Umsetzung informatischer Bildung gäbe. Das Sprachenargument besagt, dass das Programmieren eines Computers mit dem Lernen einer Sprache vergleichbar sei und argumentiert, dass es keinen Grund gäbe, jungen Menschen diese Sprache vorzuenthalten.

6 Diskussion

Betrachtet man die Art und Weise der Argumentation, so zeigt sich, dass vier Argumentationsschemata dominieren: Die Argumentation ist von praktischen Überlegungen geprägt (*Practical Reasoning*). Zudem werden Expertenmeinungen (*Argument from Position to Know*) und noch stärker allgemein akzeptierte Aussagen (*Arguments from Popular Acceptance*) als Prämissen genutzt. Eine weitere beliebte Form der Argumentation nutzt konkrete Beispiele, Präzendenzfälle oder Analogien zu anderen Fächern und Disziplinen (*Arguments Based on Cases*). In den untersuchten Dokumenten wird kaum evidenzbasiert argumentiert. Obgleich etwa das Konzeptwissensargument zusätzlich durch empirische Befunde belegt werden könnte [VI03], erfolgt dies nicht.

In den Dokumenten werden die einzelnen Argumente teilweise verknüpft. So wird etwa das Konzeptwissensargument in seiner ursprünglichen Form kaum noch verwendet. In aktuelleren untersuchten Dokumenten wird es eher als Zusatz in der Argumentation des Verständnis- und Teilhabearguments herangezogen, wobei der Fokus auf dem *mündigen* Umgang mit Informatiksystemen liegt.

Bei Betrachtung des zeitlichen Verlaufs stellt man fest, dass sich auch die Argumente gewandelt haben. Gerade in den letzten 5 Jahren hat der Begriff des Computational Thinking in den untersuchten Dokumenten an Bedeutung gewonnen. Im Unterschied zu älteren Dokumenten kommt hierbei der Übertragbarkeit informatischer Denkweisen auf Probleme außerhalb der Informatik eine größere Bedeutung zu. Generell zeigt sich in den letzten Jahren ein Wandel in den verwendeten Begrifflichkeiten: von der „Erklärung alltäglicher Informatiksysteme" hin zur „Erklärung der digitalen Welt". Während Argumente wie das Konzeptwissenargument in ihrer ursprünglichen Form also seltener verwendet werden, ist die Bedeutung anderer gestiegen: So ist das Argument, dass die persönlichen Interessen des Lernenden angesprochen werden, vor allem in aktuellen Dokumenten zu finden. Und obwohl Themen wie Simulationen als Teil der Informatik auch in den älteren Dokumenten genannt werden, findet es als explizites Wissenschaftsargument erst in aktuelleren Dokumenten Verwendung.

Eine mögliche Einschränkung der Validität dieser Untersuchung könnte die Auswahl der Dokumente darstellen, da diese vor allem den Zeitraum der letzten 25 Jahre abdecken. Da allerdings eine große Übereinstimmung mit bestehenden Kategoriensystemen besteht, kann zumindest von einer ausreichenden Repräsentativität der Auswahl ausgegangen werden.

Eine Ausweitung der Stichprobengröße sowie ihre Internationalisierung könnte weitere Einsichten liefern.

7 Fazit

Der Beitrag zeigt, dass sich ein großer Teil der im untersuchten Korpus verwendeten Argumente auf Verständnis und Teilhabe an der „digitalen Welt“ und den Arbeitsmarkt bzw. die Berufswelt konzentriert. Argumente, die auf überfachliche Kompetenzen, lernerbezogene Argumente oder den Einfluss auf die Wissenschaft abzielen, werden dagegen seltener verwendet.

Betrachtet man die zugrunde liegenden Argumentationsschemata, zeigt sich, dass vorwiegend pragmatisch argumentiert wird: Statt Vorzüge zu belegen bzw. zu begründen, wird häufig eher ausgehend von zu vermeidenden Folgen, ungeeigneten Alternativen oder drohenden Konsequenzen argumentiert. Anstelle von empirischen Belegen werden Aussagen verstärkt durch Beispiele oder Analogien gestützt. Es wird damit kaum evidenzbasiert argumentiert.

Die Ergebnisse geben Einblick in die in der Diskussion um einen verpflichtenden Informatikunterricht verwendeten Argumente, zeigen, welche Aspekte besonders herausgestellt werden, und wie die Argumentation geführt wird. Nicht zuletzt liefert der Beitrag damit auch Hinweise darauf, welche Aspekte der Informatik als allgemeinbildend wahrgenommen werden. Gleichzeitig weißt er auf Bereiche hin, in denen die (empirische) Evidenz ausgebaut werden könnte, um die Argumentation weiter zu untermauern.

Literaturverzeichnis

[An16] Andrews, Daniel Coloma: Qualitative Argumentationsanalyse als Methode der empirischen Sozialforschung. Dissertation, Universitätsbibliothek Ruhr-Universität Bochum, 2016.

[Ba07] Bayer, Klaus: Argument und Argumentation: Logische Grundlagen der Argumentationsanalyse. Vandenhoeck & Ruprecht, 2007.

[Be03] Bethge, Bernd; Drumm, Herbert; Knapp, Thomas; Neumeyer, Steffen; Romeike, Ralf; Schödel, Thomas; Wiedemann, Albert; Witten, Helmut: Informatikunterricht für alle! Ludwigsfelder Thesen. Log In, (124 S 33), 2003.

[Be17] Bergner, Nadine; Köster, Hilde; Magenheim, Johannes; Müller, Kathrin; Romeike, Ralf; Schroeder, Ulrik; Schulte, Carsten: Zieldimensionen informatischer Bildung im Elementar- und Primarbereich. Frühe informatische Bildung–Ziele und Gelingensbedingungen für den Elementar-und Primarbereich. Berlin, 2017.

[BM19] Blikstein, Paulo; Moghadam, Sepi Hejazi: Computing EducationLiterature Review and Voices from the Field. In (Fincher, Sally A.; Robins, Anthony V., Hrsg.): The Cambridge Handbook of Computing Education Research. Cambridge Handbooks in Psychology, Cambridge University Press, S. 56–78, 2019.

[Br17] Brinda, Torsten: Medienbildung und/oder informatische Bildung? DDS - Die Deutsche Schule, (2):175–187, 2017.

[DH16] Döbeli Honegger, Beat: Mehr als 0 und 1: Schule in einer digitalisierten Welt. Bern: hep Verlag, 2016.

[Fl16] Fluck, Andrew E; Webb, Mary; Cox, Margaret J; Angeli, Charoula; Malyn-Smith, Joyce; Voogt, Joke; Zagami, Jason et al.: Arguing for computer science in the school curriculum. Educational Technology & Society, 19(3):38–46, 2016.

[GF09] GFD 2009: Mindeststandards am Ende der Pflichtschulzeit: Erwartungen des Einzelnen und der Gesellschaft – Anforderungen an die Schule - Positionspapier der GFD. 2009. URL: http://www.fachdidaktik.org/cms/download.php?cat=Veröffentlichungen&file=Mindeststandards_Ende_Pflichtschulzeit.pdf.

[He96] Heymann, Hans Werner: Allgemeinbildung und Mathematik. Beltz Weinheim, 1996.

[Je10] Jeffery, Joseph: Breaking the mold: why computer science needs to be a fundamental science within the BC curriculum. In: Proceedings of the 15th Western Canadian Conference on Computing Education. ACM, S. 12, 2010.

[Ke90] Kerner, Immo O: Der Bildungskern der Informatik. In: Computer in der Schule 3, S. 189–200. Springer, 1990.

[Ma00] Mayring, Philipp: Qualitative Content Analysis. Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 1(2), 2000.

[Pa17] Passey, Don: Computer science (CS) in the compulsory education curriculum: Implications for future research. Education and Information Technologies, 22(2):421–443, 2017.

[SS15] Schauer, Carola; Schauer, Hanno: IT an allgemeinbildenden Schulen: Bildungsgegenstand und -infrastruktur. Auswertung internationaler empirischer Studien und Literaturanalyse. ICB-Research Report 63, Universität Duisburg-Essen, 2015.

[St16] Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland: , Strategie Bildung in der digitalen Welt, 2016. URL: https://www.kmk.org/aktuelles/thema-2016-bildung-in-der-digitalen-welt.html.

[VI03] Voß, Siglinde; Immenstadt, Gymnasium: Objektorientierte Modellierung von Software zur Textgestaltung. In: INFOS. S. 211–223, 2003.

[VSC17] Vogel, Sara; Santo, Rafi; Ching, Dixie: Visions of Computer Science Education: Unpacking Arguments for and Projected Impacts of CS4All Initiatives. In: Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education. SIGCSE '17, ACM, New York, NY, USA, S. 609–614, 2017.

[Wa05a] Walton, Douglas: Argumentation methods for artificial intelligence in law. Springer Science & Business Media, 2005.

[Wa05b] Walton, Douglas: Fundamentals of critical argumentation. Cambridge University Press, 2005.

[Wi03] Witten, Helmut: Allgemeinbildender Informatikunterricht? Ein neuer Blick auf HW Heymanns Aufgaben allgemeinbildender Schulen. In: INFOS. S. 59–75, 2003.

[WRM08] Walton, D.; Reed, C.; Macagno, F.: Argumentation Schemes. Cambridge University Press, 2008.

Medienkompetenz und IT-Sicherheit

Digitale Bildung und medientechnische Mündigkeit im Kontext der Wissenskluft im tertiären Bildungsbereich am Beispiel der Medienwissenschaft

Raphael Matthias Morisco[1]

Abstract: Die Forschung zur Medienkompetenz durchläuft einen stetigen Adaptionsprozess entsprechend der fachinternen Prägung der Medienpädagogik. Dabei lässt sich feststellen, dass Medienkompetenz heute schon in diversen Bereichen der Gesellschaft gefördert wird. Allerdings ist darauf hinzuweisen, dass der Aspekt der IT-Sicherheit sowie die technische Variante der Auseinandersetzung mit digitalen Medien in vielen gesellschaftlichen Gebieten und insbesondere in wissenschaftlichen Disziplinen, die sich mit den ‚Neuen Medien' beschäftigen, ausgeklammert wird oder schlichtweg fehlt. Besonders im Bereich der Medienwissenschaften bedarf es daher einer Erweiterung der Medienanalyse um den Aspekt der technischen Variante. Die hier vorliegende Arbeit erörtert die aktuelle Ungleichheit der medientechnischen Grundlagenausbildung im tertiären Bildungsbereich. Der Verlauf wird mittels des Diskurses zur Medienkompetenz mit Blick auf die fachinterne Konzeption der Medienwissenschaft dargestellt und die Notwendigkeit zur Erweiterung der Grundlagenausbildung für die Medienanalyse anhand einer didaktischen Konzeption verdeutlicht.

Keywords: Medienkompetenz, IT-Sicherheit, IT-Wissen, Wissenskorpus, Medienpädagogik, Medienanalyse, Medienbildung, tertiärer Bildungsbereich, akademische Wissenskluft, Disparität.

1 Einleitung

Bereits vor über fünfzehn Jahren wurde die Forderung im stetigen Diskurs über Medienkompetenz geäußert, IT-Sicherheit als eine verpflichtende Komponente anzusehen [ER99][Wa01]. Im Gegensatz zu den ersten Jahren nach der Jahrtausendwende wird die Medienkompetenzförderung hinsichtlich des Umgangs mit digitalen Medien inzwischen durchaus auf diversen gesellschaftlichen Ebenen gefördert [Be16], allerdings wird deutlich, dass „sich die Schule von der Digitalisierung noch immer wenig berührt [zeigt]" [Sp18]. Diese Problematik offenbart sich auch im akademischen Sektor, wobei hier die technischen Rahmenbedingungen einer digitalen Infrastruktur durchaus gegeben sind. Allerdings zeigen sich bei der Wissensvermittlung fachübergreifende Diskrepanzen. An dieser Stelle geht es nicht um die Forderung, dass IT-Wissen, im Speziellen IT-Sicherheit als Teil der Medienkompetenz, überhaupt vermittelt werden sollte, als vielmehr darum, dass geklärt werden muss, ob und wie dies überhaupt im schulischen und *besonders* im akademischen Bereich geschieht. Denn der Gedanke der Aufbereitung

[1] Universität Bielefeld, Medien- und Erziehungswissenschaft, Universitätsstraße 25, 33615 Bielefeld, r.morisco@uni-bielefeld.de

doi:10.18420/inf2019_78

eines Basiswissens über IT-Sicherheit als fester Bestandteil von Medienkompetenz, ist fest verbunden mit der Diskussion zur soziokulturellen Problematik der Wissenskluft sowie digitalen Kluft [Bo94] [Bo04] [St02]. Der Zugang zu den jeweiligen Medien- und Technologiesystemen ist vereinzelt noch heterogen und eine divergente Wissensaufnahme führt zu einer neuen Wissenskluft mit gravierenden Auswirkungen im Zuge der Digitalisierung. Hierbei ist der Fokus nicht drauf gerichtet, dass die Wissenselite von den neuen digitalen Medien mehr profitiert als das Wissensprekariat, sondern es wird die technische Variante der digitalisierten Medien als Beherrschungs- und Reflexionswissen in vielen Bereichen systematisch ausgeklammert. Auf akademischer Ebene wird dieser Aspekt durch einen auffallenden Mangel bezüglich des Inhalts der Studienpläne der Medienwissenschaften an deutschen Universitäten deutlich, wie eine Querschnittsanalyse zur medienwissenschaftlichen Lehre an zehn arbiträr ausgewählten Universitäten zeigt [Mo16:2f, 95ff]. Konkret bedeutet dies, dass die mangelnde Betrachtung der IT-Sicherheitsthematik in Bezug auf das Erlernen der Medienanalyse von Medienwissenschaftlern, insbesondere hinsichtlich digitaler Medien, ein aufkommender Katalysator für eine Wissenskluft im akademischen Bereich ist, die sich auf die Gesellschaft ausweiten wird.

Ziel des vorliegenden Beitrages ist es, die aktuelle Ungleichheit der medientechnischen Grundausbildung im tertiären Bildungsbereich anhand des Diskurses zur Medienkompetenz mit Blick auf die fachinterne Konzeption der Medienwissenschaft zu analysieren. Im Zuge dessen wird die Notwendigkeit zur interdisziplinären Erweiterung der Grundlagenausbildung für die Medienanalyse durch didaktische Ansätze verdeutlicht. In Kapitel 2 wird zunächst die Ausgangssituation hinsichtlich der kritischen Begriffsgenese und der damit verbundenen Diskursfelder erläutert und die Kategorisierung der Risiko- und Datenschutzkompetenz in der Diskussion dargestellt. Darüber hinaus werden die Herausforderungen von Big Data und einer technologiezentrierten Medienkompetenz verdeutlicht. In Kapitel 3 wird die Problematik der curricularen Auseinandersetzung mit IT-Sicherheit im akademischen Bereich am Beispiel der Medienwissenschaft erörtert. Im Fokus liegt dabei die Neuausrichtung der Medienanalyse. Der Beitrag schließt in Kapitel 4 mit einer offenen Diskussion über eine skizzierte Konzeption für die didaktisch Umsetzung der Medienanalyse in der Medienwissenschaft in Hinblick auf die Interdisziplinarität.

2 Medienkompetenz im interdisziplinären Diskurs

Der Begriff *Medienkompetenz* kann als beständiges Konzept verstanden werden, das in sozialen Diskursen auch in der gegenwärtigen Zeit äußerst populär ist. Es zeigt sich, dass die Bereiche Politik, Wirtschaft sowie einzelne wissenschaftliche Disziplinen den Begriff immer wieder als zentrales Konzept unserer Zeit betrachten und entsprechend nutzen [GOS17a] [JW09] [Sc05]. Mit Blick auf den aktuellen Diskurs und die Konjunktur zur Begriffsbedeutung kann festgestellt werden, dass die Auseinandersetzung vollständig von den Veränderungsprozessen bestimmt wird, die durch die Medientechnologien in

unserer Gesellschaft ausgelöst werden [GOS17b]. Hierbei wird deutlich, dass die Zielbestimmung von Medienkompetenz, „die Fähigkeit, mit Medien im technologischen wie auch sozialen Bezug und im Hinblick auf persönliche Entwicklungsziele erfolgreich umgehen zu können“ [St02] und die Forderung danach, nicht obsolet sondern von bedeutender Aktualität sind. Allerdings muss angemerkt werden, dass unterschiedliche Diskurse das konzeptionelle Begriffsverständnis formen und je nach Perspektive des Akteurs unterschiedliche Aspekte und Charakteristika betont werden.

2.1 Begriffsgenese und Diskursfelder

Nach Gaspki, Oberle und Staufer können sechs Diskurse skizziert werden, die die Deutung und Auslegung von Medienkompetenz in ihren unterschiedlichen Formen prägen: Der wirtschaftliche, der rechtliche, der medienpolitische, der gesellschaftspolitische, der medientechnische und der pädagogische Diskurs [GOS17b: 19ff]. Der Terminus selbst ist „ein massenmediales Konstrukt, dessen Konjunktur im gegenwärtigen Strukturwandel der modernen funktional differenzierten Medien- und Informationsgesellschaft begründet liegt“ [Ga17: 110]. Dabei folgt die Forderung nach neuen Kompetenzen im bildungspolitischen Diskurs den jeweiligen medientechnologischen Innovationen, wie dies exemplarisch durch die fachinterne Entwicklung der Medienpädagogik gezeigt werden kann. Insofern „ist die Geschichte der Medienpädagogik eine Geschichte der Reaktion auf die jeweils ‚neuen Medien‘ und die durch sie hervorgerufenen Irritationen“ [HP05]. In der medienpädagogischen Forschung sieht dieser Analyseverlauf wie folgt aus: Medien, verstanden als Sozialisationsinstanzen [Sü13] [Au08], werden retrospektiv analysiert. Darauf aufbauend werden Prognosen bezüglich bildungsspezifischer Prozesse abgeleitet und zudem Handlungsempfehlungen angeführt, wie die entsprechende Medientechnik förderlich und nutzbringend eingesetzt werden kann. Dabei werden neben rein deskriptiven Ansätzen vor allem qualitative und quantitative Methoden der empirischen Sozialforschung angewendet. Die gesellschaftlichen Veränderungen und die damit verbundenen Diskurse können in der Forschung der Medienpädagogik als ein sich stetig transformierendes Abbild des Mediensozialisationsprozesses in der Gesellschaft dargestellt werden.

Mit Blick auf die bildungspolitische Diskussion besteht die Problematik darin, dass die Medienbildung in dieser Auseinandersetzung eine Sonderstellung einnimmt, „weil es sich bei ihr nicht um ein Unterrichtsfach, sondern um einen fächerübergreifenden Bereich handelt“ [Tu10: 81]. Für die schulische Ausrichtung einer Medienkompetenzförderung lassen sich zahlreiche Ansätze finden und anführen, die sich dabei stets in die oben angeführten sechs Diskurse entsprechend eingliedern lassen [Pi18] [Ma17] [Hu15]. Dabei offenbart sich, dass in diversen Publikationen semantische Überschneidungen von Termini mit dem Konzept der Medienkompetenz gegeben sind, vornehmlich auf Ebene der politischen Bildung mit dem Schwerpunkt Medienbildung [GOS17b: 19] [SG13: 103f] [SG07: 281f]. Exemplarisch zeigt dies der Begriff der „Digitalen Souveränität“ [BN14: 763] im Kontext der Thematik der IT-Sicherheit. In der akademischen Ausrichtung der Medienkompetenzförderung wird deutlich, dass es ebenfalls

Besterbungen gibt eine eigene „akademische Medienkompetenz“ im Diskurs an Hochschulen zu etablieren [RHF14] [KH19: 157f], allerdings ist der analytische Schwerpunkt der Medienpädagogik auf die Umsetzung von Implementierungsprozessen diverser IT-Systeme für den akademischen Bereich gerichtet, um die Mediekompetenzen von Lehrenden und Studierenden entsprechend zu verbessern und zu erweitern [Ma17] [NS15].

2.2 Risiko- und Datenschutzkompetenz

Darüber hinaus lassen sich in der medienpädagogischen Forschung zwei Kompetenzbeschreibungen als Bestandteile von Medienkompetenz aufzeigen, die in den letzten Jahren im medientechnischen Diskurs wieder an Bedeutung gewonnen haben: die Risiko- und die Datenschutzkompetenz. Nach Klebl und Brost ermöglicht die Risikokompetenz es dem Nutzer „Ungewissheit reflexiv zu erkennen und entweder in Fachwissen zu überführen oder durch Rückgriff auf Heuristiken zu bewältigen“ [KB10: 241]. Hierbei ist Ungewissheit als wissensspezifische Unsicherheit zu verstehen. Diese Problematik hat in Zeiten von informationstechnologischen Systemen und Algorithmen in Form von ‚bots‘ und ‚social bots‘ insbesondere im Kontext des Journalismus im 21. Jahrhundert und der politischen Meinungsbildung an gesellschaftlicher Brisanz rasant zugenommen [Sc17a] [NLN14]. Infolgedessen rückt die Forderung nach einer Risikokompetenz als ergänzender Bestandteil der Medienkompetenz wieder in den medienpädagogischen Fokus. Des Weiteren kristallisieren sich im Diskurs verstärkt die Forderungen nach Bestandteilen der Medienkompetenz heraus, wie etwa Privatsphäre, IT-Sicherheit, informationeller Selbstbestimmung und Anonymität. Gimmler fordert, mit Bezug auf das eigene Medienkompetenzkonzept für die Schule [SG07], dass Datenschutzkompetenz als elementarer Bestandteil von Medienkompetenz zu vermitteln ist [Gi12: 111f]. Dieser Forderung folgend beschreiben Hug und Grimm ein Datenschutzkompetenzmodell [HG17: 169], indem sie auf das Medienkompetenzmodell nach Six und Gimmler Bezug nehmen, allerdings ohne die Punkte der Risikobewertung und Vermeidungsstrategie. Diese werden von den Autoren entsprechend zu einem erweiternden Datenschutzkompetenzmodell hinzugefügt.

2.3 Die Herausforderungen: Big Data und das deutsche Bildungssystem

Infolgedessen ist in Hinblick auf die oben angeführten Kernattribute eine Priorisierung in allen Diskursen zur Medienkompetenz auf den Datenschutz als Bildungsaufgabe festzustellen [GOS17b]. Der Fokus liegt dabei hauptsächlich auf der Konzeption und Ausrichtung einer Didaktik für die schulische Bildung. Diese Priorsierung ist im Forschungsfeld der Medienpädagogik auf einen erneuten Adaptionsprozess zurückzuführen, indem durch die progressive medientechnologische Entwicklung abermals ein Perspektivwechsel für die Orientierung des wissenschaftlichen Gegenstandes im Diskurs entstanden ist. Demzufolge wurde, vor allem nach den Enthüllungen der globalen Überwachungs- und Spionageaffäre der Blick, angesichts der

Konvergenzprozesse der Digitalisierung in der Gesellschaft vermehrt auf die ökonomische Verwendung von Daten durch internationale Großunternehmen wie Google, Facebook, Microsoft und Apple gerichtet. Hierbei geht es wesentlich um die Analyse der Transformation hinsichtlich des Umgangs mit Informationen. In diesem Zusammenhang lautet das zentrale Schlagwort in den wissenschaftlichen Diskursfeldern: Big Data. Historiographisch betrachtet, bedeutet das, wenn bis vor etwa zehn Jahren das individuelle Datum im Mittelpunkt des Interesses von Unternehmen stand, so „änderte sich die Perspektive mit den riesigen, jederzeit verfügbaren Datenmengen dramatisch. ‚Big Data' steht wie kein anderer Begriff für den Übergang zu einem neuen Modell des Umganges mit Informationen" [Sc17b: 75] In Anlehnung an Gapski ist der Thematik im Kontext einer medienpädagogischen Auseinandersetzung ein bedeutender gesellschaftlicher und wissenschaftlicher Stellenwert zuzuschreiben, insofern „Big Data zunehmend zu einem lebensweltlich erfahrbaren Phänomen [wird], wenn Konsumtenentscheidungen, Informationsverhalten und Karrierechancen in den Anwendungsbereich algorithmisierter Prozesse fallen" [Ga15b: 13].

2.4 Technologiezentrierte Medienkompetenz als medienpädagogischer Fokus

Unter Beachtung all dieser Aspekte in der Diskussion um Medienkompetenzmodelle kann resümierend bezüglich des medientechnischen Diskurses festgestellt werden, dass sich den angeführten Ausarbeitungen über die Modelle und Konzepte von Medienkompetenz und den verschiedenen Erweiterungsformen stets die Ableitung möglicher Handlungsempfehlungen im Bereich der Medienbildung anschließt [GOS17b] [HG17]. Dabei zeigt sich, dass im Kontext der Konzeption einer technologiezentrierten Medienkompetenz die Forderung nach einer IT-Sicherheitsdidaktik im Diskurs der letzten zwanzig Jahre wiederholt gestellt wird. Das heißt, alle medientechnischen Kompetenzerweiterungen haben eine zentrale Forderung gemein, die ein Bewusstsein für informationstechnologische Sicherheit als Teil von Medienkompetenz zwingend macht [GOS17a] [He16]. Folglich muss diese Form der Kompetenz hinsichtlich der Technologie „sowohl ein informationstechnisches Grundverständnis von Aufbau und Funktionsweise der Informations- und Kommunikationstechnologien umfassen als auch die Fähigkeiten sich der Datenschutztechniken zu bedienen" [Wa01: 13]. Ebendiese Aspekte werden in der Diskussion über *Informatik für alle* mit dem Schwerpunkt auf dem deutschen Schulsystem in einer allgemeineren Form angesprochen und entsprechend diskutiert [La18a] [La18b]. Das heißt, das gesamte Fach Informatik steht dabei im Fokus und nicht nur ausgewählte Bereiche.

Die Problematik, die Lautebach im Diskurs zur digitalen Mündigkeit hierbei verdeutlicht, spiegelt die konkrete fachinterne Herausfoderung der Medienwissenschaften im 21. Jahrhundert wider. Nutzer digitaler Medien können die Systeme zwar benutzen, jedoch können sie diese „weder verstehen noch hinterfragen, noch kompetent auswählen und schon gar nicht nach eigenen Vorstellungen anpassen" [La18a]. Daran anknüpfend können sich Nutzer in der Informationsgesellschaft keine Meinung zu technisch geprägten Themen bilden. Dass diese Darstellung für den

akademischen Bereich etwas überspitzt wirken mag, lässt sich damit begründen, dass es gewiss ein informatisches Bildungsangebot fachübergreifend für Studierende an deutschen Hochschulen gibt, wenngleich die eigentliche Problematik fachintern in den Medienwissenschaften bestehen bleibt. Denn der Diskurs über IT-Sicherheit als Bestandtteil der Medienkompetenz ist insgesamt im angeführten Fachbereich fragmentarisch bis inexistent [Mo16]. Dies ist zum einen auf die inhärent disziplinäre Ausrichtung der drei Hauptarbeitsbereiche Mediengeschichte, Medientheorie und Medienanalyse zurückzuführen, wobei die Konzeptualisierung in Anlehnung an die von den älteren Wissenschaften entwickelten Verfahren, Begriffssysteme und Modellen erfolgt. Zum anderen ist dies laut Herzig auf die Gewichtung der Medienpädagogik zurückzuführen, bei der „die Thematisierung informatorischer Aspekte […] eher die Ausnahme bilden" [He16: 59]. Folglich geht es primär um die mediendidaktische Analyse der Anwendung von informations- und kommunikationstechnologischen Systemen im Hochschulbereich. Dies wird unter anderem durch die verschiedenen Themenfelder der Hochschulforschung bestätigt [Wo15: 153f]. Mit Blick auf Hochschulen und Universitäten wird damit deutlich, dass im Akademischen die Auseinandersetzung mit IT-Sicherheit vornehmlich in den jeweiligen technischen Fachgebieten stattfindet [Ec14].

Für den Fachbereich der Medienwissenschaft – im Speziellen für den Arbeitsbereich der Medienanalyse – wird die medientechnische Untersuchung und die damit verbundene kritische Auseinandersetzung zusehends bedeutsamer. Es stellt sich angesichts algorithmischer Prozesse, der Kommunikationsverfahren im Zuge der Digitalisierung und exemplarisch anhand der sozialen Netzwerke im Internet die unvermeidbare Frage: „Echt oder nicht, Mensch oder Maschine?" [Sa17]. Das heißt, gemäß Salewski, „richtig sicher können wir uns bei keinem Post, Tweet oder Kommentar mehr sein" [Sa17], wer eigentlich mit wem kommuniziert. Dabei wird vor allem mit Blick auf die Online-Kommunikation deutlich, dass die Grenzen der interpersonalen Kommunikation für das Individuum durch den wachsenden Einfluss informationstechnologischer Artefakte[2] nicht mehr klar erkennbar sind. In diesem Punkt offenbart sich zum einen das grundlegende Problem der Forschung zur Medienwissenschaft und zum anderen die Notwendigkeit einer zeitgemäßen medientechnischen Analyse, die sich aktuell noch auf die reinen Bedeutungskonstruktionen, Strukturen und den ästhetischen Gehalt von Medienprodukten beschränkt. Eine Untersuchung zur Weiterentwicklung der Medienwissenschaft durch eine Neuausrichtung und adäquate sowie sachbezogene Schwerpunktsetzung im Bereich der „sektorialen Gliederung der Medienanalyse" [Hi10: 339] liegt bislang nicht vor.

[2] Gemäß der semantischen Beschreibung, etwas Von-Menschenhand-Geschaffenes, sind darunter im gegebenen Kontext u. a. einfache bis komplexe Algorithmen, Bots und Social Bots zu verstehen.

3 Der Wissenskorpus und die Neuausrichtung der Medienanalyse

Im Zuge der voranschreitenden Digitalisierung und vor allem im Kontext der oben angeführten Wissenskluft ist die Problematik der mangelhaften Medienkompetenz von Studierenden der Medienwissenschaften nur durch die Aufbereitung eines Wissenskorpus über IT-Sicherheit zu lösen, um so eine Verbesserung der Medienanalyse zu ermöglichen.[3] In diesem Zusammenhang sollten auch die technischen Aspekte im Sinne eines Basiswissens berücksichtigt werden.

3.1 Das Basiswissen über IT-Sicherheit und die fachspezifische Anforderung für die Medienwissenschaft

Der geforderte Wissenskorpus lässt sich als aufbereitetes Basiswissen zur IT-Sicherheit darstellen, das sich in zwei wesentliche Bereiche gliedern lässt. Zum einen in den Bereich der grundlegenden IT-Sicherheitsbegriffe und zum anderen in den Bereich der Bedrohungen für IT-Systeme. Dabei ist zu beachten, dass Bedrohungen für IT-Systeme immer in Verbindung mit möglichen Lösungsansätzen sowie präventiven Maßnahmen betrachtet werden müssen. Unter dem Terminus IT-Sicherheit können gemäß Eckert folgenden Begriffe beziehungsweise Eigenschaften subsumiert werden: Sicherheit (mit den Untereigenschaften: Funktions-, Informations-, Datensicherheit, Datenschutz), Verlässlichkeit, Authentizität, Datenintegrität, Informationsvertraulichkeit, Verfügbarkeit, Verbindlichkeit und Anonymisierung [Ec14: 6ff].

Die Problematik von IT-Systemen und der daraus resultierende Stellenwert für den medienwissenschaftlichen Diskurs werden größtenteils durch die Informationsflut in den Massenmedien deutlich. Dies zeigt sich exemplarisch durch den medialen Umgang mit der informationstechnologischen Thematik des Hackens und der daraus resultierenden medialen Darstellung. Hierbei tritt insbesondere die Sozialfigur des Hackers in den Vordergrund, der in der medialen Aufbereitung überwiegend negativ konnotiert ist [Mo16: 9ff]. Der Hacker als Sozialfigur und das Hacken als potenzielle individuelle und gesellschaftliche Gefahrenquellen sind im medienwissenschaftlichen Diskurs über neue Medien insofern von elementarer Bedeutung, als sie einen Umgang mit der Technik darstellen, der über reines Anwendungswissen hinausgeht und deren Wissensvorsprung – je nach Handlungsabsicht – für den Nutzer gefährlich werden kann. Dabei wird bei näherer Betrachtung deutlich, dass die Gefühle von Bedrohtsein und Gefahr lediglich aus mangelndem Wissen über digitale Systeme resultieren und sich jeder einfache Nutzer vor möglicher Ausspähung oder Datendiebstahl schützen könnte, wenn er über eine fundierte Medienkompetenz im Bereich Digitale Medien verfügte. In diesem Zusammenhang ist insbesondere die Reflexion von Basiswissen ein kritischer Punkt, um zu verdeutlichen,

[3] Unter dem Begriff ‚Korpus' wird „eine Sammlung von (repräsentativen) Texten […] zum Zwecke wissenschaftlicher Untersuchungen" [Cy15] verstanden. Das heißt, in Anlehnung an die Literaturwissenschaften, umfasst der hier konzipierte Begriff des Wissenskorpus repräsentative Texte und Ausarbeitungen zur IT-Sicherheitsthematik.

dass sich die Komplexität von technischen Entwicklungen durch ihren inhärenten Aufbau ohne Basiswissen dem Individuum verschließen. Exakt dieser Aspekt ist es, der die Wissenskluft schafft und eine Kluft zwischen den Usern und dem Bereich der Hacker forciert. In Kombination mit der beschriebenen Medienwirklichkeit ergibt sich die fachspezifische Anforderung für die Neuausrichtung der Medienanalyse von neuen Medien.

3.2 Interdisziplinarität als Nutzen für die fachinterne Weiterentwicklung

Der Nutzen, der für Studierende entsteht, ist das Anstoßen eines Bewusstseins für die kritischen Aspekte von IT-Systemen, die sich im Zuge der Digitalisierung immer weiter in der Gesellschaft verankern werden. An dieser Stelle geht es nicht um den Gesichtspunkt der Emotionalisierung und das Schüren von Ängsten im Hinblick auf die Bedrohungen und Angriffe, denen IT-Systeme ausgesetzt sind. Vielmehr geht es um die Gefühle von Bedrohtsein im Informationszeitalter, die nur aus einem mangelnden Wissen über digitale Systeme selbst heraus entstehen können. Dieser fundamentale Aspekt muss an Studierende weitergegeben werden, wenn eine evolvierende Mündigkeit, beginnend auf akademischer Ebene, Teil der Entwicklung des Nutzers werden soll. Nur durch die Eingliederung der IT-Sicherheitsthematik in den wissenschaftlichen Diskurs über Medien kann die bestehende Wissenskluft auch fachintern aufgelöst werden und so ein interdisziplinärer Diskurs über neuen Medien ermöglicht werden, der die medientechnischen Aspekte adäquat integriert. Außerdem wird eine fachinterne Auseinandersetzung mit den Unterpunkten der informationellen Selbstbestimmung, Privatsphäre und informationstechnologischen Sicherheit auf medientechnischer Ebene möglich. Zudem betrifft dies den Journalismus als Teil der medienwissenschaftlichen Lehre, in welchem das Wissen über IT-Sicherheit für angehende Journalisten von existenzieller Bedeutung ist.

4 Diskussion: Didaktische Konzeption für die Medienanalyse

Im Folgenden und abschließend soll eine kurze didaktische Konzeption zur Diskussion gestellt werden. Hierbei dient der geforderten Wissenskorpus mit dem Basiswissen über IT-Sicherheit als Ausgangspunkt, um eine mögliche medienpädagogische Herangehensweise zu skizzieren.

In der fachinternen Konzeption der Medienwissenschaften lassen sich allgemein zwei Hauptrichtungen beziehungsweise Verfahrensweisen unterscheiden. Zum einen die historisch-hermeneutische und zum anderen die empirisch-analytische Ausrichtung. Im Hinblick auf eine Vorlesung oder ein Seminar wäre somit eine gangbare Gliederung für eine medienwissenschaftliche Veranstaltung, die beide Verfahrensweisen kombiniert, beispielhaft wie folgt zu konzipieren: Ausgehend von der transmedialen Darstellung des Hackers, über dessen Fähigkeiten mit Blick auf die Mediensysteme, bis hin zur Fragestellung, wie man sich vor informationstechnologischen Bedrohungen effektiv schützen kann, wird erarbeitet, welches Wissen Studierende dafür benötigen, um die medientechnische Ebene abschätzen und fachspezifisch auslegen zu können. Hierbei kann die

transmediale Darstellung der Sozialfigur des Hackers als thematischer Einstieg verstanden werden, der durch ausgewählte Artikel aus Online- und Printmedien ermöglicht wird. Die Studierenden sollten dann exemplarisch die Fragen klären welche Eigenschaften von IT-Sicherheit bei den ausgewählten Artikeln genau bestimmt werden können und welche Problemstellungen sich medientechnisch kategorisieren lassen könnten. Eine zentrale Frage könnte dabei lauten: Wurde eine der vier Untereigenschaften des Bereichs der Sicherheit verletzt? Dabei ist anzumerken, auch wenn sich die Grenzen zwischen Funktions- und Informationssicherheit durchaus überschneiden können [Ec14: 7], sollten die Studierenden die Grundbegriffe an faktischen Beispielen lernen. Gehen wir im Verlauf der skizzierten Veranstaltung davon aus, dass exemplarisch Texte gegeben sind, die unter anderem die Eigenschaft von Datenschutz kompromittiert zeigen. Dieser Aspekt würde es ermöglichen die technischen Möglichkeiten von Anonymisierung sowie Verschlüsselung von Daten kontextbezogen zu integrieren und die Studierenden in einer Blended-Learning-Veranstaltung[4] einzubinden. Dabei würden Studierenden mit Anonymisierungstools vertraut gemacht und könnten medientechnisch die Vor- und Nachteile der digitalen Werkzeuge in einem medienwissenschaftlichen Kontext im Hinblick auf Verständlichkeit und Anwenderfreundlichkeit diskutieren. Darüber hinaus könnten entsprechend die Vorteile des E-Learnings mit denen einer Präsenzveranstaltung kombiniert werden, um das benötigte Basiswissen zu vermitteln. Besonders Studierenden, die nach Abschluss ihres medienwissenschaftlichen Studiums ihren Weg weiter im Bereich des Journalismus gehen möchten, könnte so ein elementarer Aspekt nahegebracht werden, der für den modernen Journalismus unumgänglich ist: anonymisierte Kommunikation.

Literaturverzeichnis

[Au08] Aufenanger, S.: Mediensozialisation. In (Sander, U.; Gross, F.; Hugger, K.-U., Hrsg.): Handbuch Medienpädagogik. VS Verlag für Sozialwissenschaften, GWV Fachverlage GmbH, Wiesbaden, S. 87–92, 2008.

[Be16] Becker, L.: Digitalisierung in der Schule. Lernen für eine neue Welt, https://www.faz.net/aktuell/beruf-chance/campus/digitalisierung-in-der-schule-lernen-fuer-eine-neue-welt-14534122.html, 19.11. 2016. Stand: 10.04.2019.

[BN14] Boberach, M.; Neuburger, R.: Zukunftspfade Digitales Deutschland 2020. HMD Praxis der Wirtschaftsinformatik, Dezember, S. 762–772, 2014.

[Bo04] Bonfadelli, H. Medienwirkungsforschung 1: Grundlagen und theoretische Perspektiven. UTB, Stuttgart, 2004.

[Bo94] Bonfadelli, H.: Die Wissenskluftperspektive. Massenmedien und gesellschaftliche Information. UVK-Medien, Konstanz, 1994.

[4] Unter Blended-Learning versteht man die Kombination von unterschiedlichen Methoden und Medien, an der Universität etwa aus Präsenzveranstaltung und E-Learning.

[Cy15] Cyffka, A.: PONS. Großwörterbuch Deutsch als Fremdsprache. Pons, Stuttgart, S. 832, 2015.

[Ec14] Eckert, C.: IT-Sicherheit. Konzepte – Verfahren – Protokolle. De Gruyter, München, 2014.

[ER99] Espey, J.; Rudinger, G.: Der überforderte Techniknutzer. IT-Sicherheit aus psychologischer Sicht. Praxis der Informationsverarbeitung und Kommunikation (PIK), S. 178–185, 1999.

[Ga15] Gapski, H.: Big Data und Medienbildung – eine Einleitung. In (Gapski, H, Hrsg): Big Data und Medienbildung. Zwischen Kontrollverlust, Selbstverteidigung und Souveränität in der digitalen Welt. kopaed verlagsGmbH, Düsseldorf, u.a., S. 9–18, 2015.

[Ga17] Gapski, H.: Politisch orientierte Medienkompetenzförderung inmitten der digitalen Transformation. In (Gapski, H.; Oberle, M.; Staufer, W., Hrsg): Medienkompetenz. Herausforderung für Politik, politische Bildung und Medienbildung. Bundeszentrale für politische Bildung, Bonn, S. 105–115, 2017.

[Gi12] Gimmler, R.: Medienkompetenz und Datenschutzkompetenz in der Schule. Datenschutz und Datensicherheit – DuD, Februar, S. 110–116, 2012.

[GOS17a] Gapski, H.; Oberle, M.; Staufer, W.: Medienkompetenz. Herausforderung für Politik, politische Bildung und Medienbildung. (Gapski, H.; Oberle, M.; Staufer, W., Hrsg.). Bundeszentrale für politische Bildung, Bonn, 2017a.

[GOS17b] Gapski, H.; Oberle, M.; Staufer, W.: Einleitung. In (Gapski, H.; Oberle, M.; Staufer, W., Hrsg): Medienkompetenz. Herausforderung für Politik, politische Bildung und Medienbildung. Bundeszentrale für politische Bildung, Bonn, S. 17–30, 2017b.

[He16] Herzig, B.: Medienbildung und Informatische Bildung – Interdisziplinäre Spurensuche. MedienPädagogik, S. 59–79, 28.10.2016.

[HG17] Hug, A.; Grimm, R.: Entwicklung eines Datenschutzkompetenzmodells. In (Diethelm, I., Hrsg.): Informatische Bildung zum Verstehen und Gestalten der digitalen Welt, Lecture Notes in Informatics, Gesellschaft für Informatik, Bonn, S. 167–170, 2017.

[Hi10] Hickethier, K.: Einführung in die Medienwissenschaft. Verlag J. B. Metzler, Stuttgart, u.a., 2010.

[HP05] Hüther, J.; Podehl, B.: Geschichte der Medienpädagogik. In (Hüther, J.; Schrob, B., Hrsg.): Grundbegriffe Medienpädagogik. kopaed, München, S. 116–127, 2005.

[Hu15] Hugger, K.-U.; et al: Jahrbuch Medienpädagogik 12. Kinder und Kindheit in der digitalen Kultur. Springer VS, Wiesbaden, 2015.

[JW09] Jarren, O.; Wassmer, C.: Medienkompetenz – Begriffsanalyse und Modell. Ein Diskussionsbeitrag zum Stand der Medienkompetenzforschung. Medien und Erziehung, S. 46–51, 2009.

[KB10] Klebl, M.; Brost, T.: Risikokompetenz als Teil der Medienkompetenz – Wissensformen im Web 2.0. In (Herzig, B.; et al., Hrsg.): Jahrbuch Medienpädagogik 8. Medien-

kompetenz und Web 2.0. VS Verlag für Sozialwissenschaften, Wiesbaden, S. 239–254, 2010.

[KH19] Kergel, D.; Heidkamp-Kergel, B.: Der kritische Dialog – Überlegungen zur akademischen Medienkompetenz im digitalen Zeitalter. In (Jahn, D.; Kenner, A.; Kergel, D.; Heidkamp-Kergel, B., Hrsg.). Kritische Hochschullehre. Diversität und Bildung im digitalen Zeitalter. Springer VS, Wiesbaden, S. 153–162, 2019.

[La18a] Lautebach, U: Informatik für alle! Ein Plädoyer, https://gi.de/meldung/informatik-fuer-alle-ein-plaedoyer, 01.02.2018. Stand: 16.06.2019.

[La18b] Lautebach, U: Informatik für alle, https://www.zeit.de/gesellschaft/schule/2018-02/digitalisierung-informatikunterricht-schulen-bildung, 21.02.2018. Stand: 16.06.2019.

[Ma17] Mayrberger, K.; et al.: Jahrbuch Medienpädagogik 13. Vernetzt und entgrenzt – Gestaltung von Lernumgebungen mit digitalen Medien. Springer VS, Wiesbaden, 2017.

[Mo16] Morisco, R.: Hacken, Cracken, Sniffen. Erweiterung der Medienkompetenz mit Blick auf die IT-Sicherheit im Digitalen Zeitalter. Masterarbeit. Bielefeld, 2016.

[NLN14] Neuberger, C.; Langenohl, S.; Nuernbergk, C.: Social Media und Journalismus. LfM-Dokumentation, Düsseldorf, 2014.

[NS15] Nistor, N.; Schirlitz, S.: Digitale Medien und Interdisziplinarität. Herausforderungen, Erfahrungen, Perspektiven. Waxmann, Münster, u.a., 2015.

[RHF14] Reinmann, G.; Hartung, S.; Florian, A.: Akademische Medienkompetenz im Schnittfeld von Lehren, Lernen, Forschen und Verwalten. In (Imort, P.; Niesyto, H., Hrsg.): Grundbildung Medien in pädagogischen Studiengängen. Schriftenreihe Medienpädagogik interdisziplinär, kopaed, München, S. 319–332, 2014.

[Sa17] Salewski, S.: Social Bots - gefährlich oder nur nervig? https://www.deutschlandfunknova.de/beitrag/social-media-sind-social-bots-maechtig-oder-nur-nervig, 7.10.2017. Stand: 28.03.2019.

[Sc05] Schorb, B.: Medienkompetenz. In (Hüther, J.; Schrob, B., Hrsg.): Grundbegriffe Medienpädagogik. kopaed, München, S. 257–262, 2005.

[Sc17a] Schweiger, W.: Der (des)informierte Bürger im Netz. Wie soziale Medien die Meinungsbildung verändern. Springer, Wiesbaden, 2017.

[Sc17b] Schaar, P.; Überwachung, Algorithmen und Selbstbestimmung. In (Gapski, H.; Oberle, M.; Staufer, W., Hrsg): Medienkompetenz. Herausforderung für Politik, politische Bildung und Medienbildung. Bundeszentrale für politische Bildung, Bonn, S. 73–81, 2017.

[SG07] Six, U.; Gimmler, Roland: „Kommunikationskompetenz, Medienkompetenz und Medienpädagogik.“ In (Six, U.; Gleich, U.; Gimmler, R., Hrsg.): Kommunikationspsychologie und Medienpsychologie. Beltz Verlag, Weinheim, u.a., S. 271–296, 2007.

[SG13] Six, U.; Gimmler, R.: „Medienkompetenz im schulischen Kontext.“ In (Vogel, I. C., Hrsg.): Kommunikation in der Schule. Klinkhardt, Bad Heilbrunn, 96–117, 2013.

[Sp18] Spiewak, M.: Fehler 404. Die Zeit 6, S. 2, 2018.

[St02] Stötzel, B.: Medienkompetenz. In (Schanze, H., Hrsg.): Metzler-Lexikon Medientheorie – Medienwissenschaft. Ansätze – Personen – Grundbegriffe. Verlag J. B. Metzler, Stuttgart, u.a., S. 225–226, 2002.

[Sü13] Süss, D.: Sozialisation. In (Bentele, G.; Brosius, H.-B.; Jarren, O.): Lexikon Kommunikations- und Medienwissenschaft. Springer VS, Wiesbaden, S. 321, 2013.

[Tu10] Tulodziecki, G.: Standards für die Medienbildung als eine Grundlage für die empirische Erfassung von Medienkompetenz-Niveaus. In (Herzig, B.; et al., Hrsg.): Jahrbuch Medienpädagogik 8. Medienkompetenz und Web 2.0. VS Verlag für Sozialwissenschaften, Wiesbaden, S. 81–101; 2010.

[Wa01] Wagner, W.-R..: Datenschutz, Selbstschutz, Medienkompetenz: Wie viel Informationstechnische Grundbildung Braucht Der Kompetente Mediennutzer? MedienPädagogik: Zeitschrift Für Theorie Und Praxis Der Medienbildung, S. 1–16, 2001.

[Wo15] Wolter, A.: Hochschulforschung. In (Reinders, H., Hrsg.): Empirische Bildungsforschung. Gegenstandsbereiche. Springer VS, Wiesbaden, S. 149–164, 2015.

Informatische Bildung als Verbraucherschutz für reflektierte Handlungen in der digitalen Welt

Schulungsmöglichkeiten für eine reflektierte Nutzung eines Smartphones

Manuel Froitzheim[1], Michael Schuhen[2] und Timo Stentenbach[3]

Abstract: Im Rahmen des Aufsatzes wird die Entwicklung und Evaluierung einer Smartphone-Simulation dargestellt, mit deren Hilfe Schülerinnen und Schüler für Handlungen in der digitalen Welt sensibilisiert werden können. Im Fokus steht ein reflektierter Umgang mit personenbezogenen Daten, wobei im Rahmen der Simulation die Handlungen der Lernenden analysiert werden und anschließend diskutiert. Durch dieses Vorgehen werden die realen Handlungen in den Mittelpunkt gestellt und nicht nur das sozial gewünschte Verhalten, dass ansonsten sehr oft von Anwendern geäußert wird, reflektiert.

Keywords: Verbraucherschutz, Digitale Bildung, elektronisches Schulbuch

1 Einleitung

Die JIM-Studie zum Medienumgang von Jugendlichen zwischen 12 und 19 Jahren im Jahr 2017 zeigt, dass die Nutzung von mit dem Internet verbundenen Anwendungen für Jugendliche bereits selbstverständlich geworden ist. 99 Prozent der Jugendlichen sind mindestens selten online und 89 Prozent sind sogar täglich online [FPR17]. Im Durchschnitt schätzen die Jugendlichen ihre tägliche Nutzung des Internets auf 221 Minuten pro Tag ein [FPR17] und 81 Prozent der Jugendlichen gibt an, dafür ein Smartphone zu verwenden [FPR17].

Innerhalb dieser täglichen 221 Minuten, aber auch während das Smartphone sich ausgeschaltet auf dem Tisch oder in der Tasche befindet, produzieren die mit Kameras, Mikrofonen und verschiedensten weiteren Sensoren ausgestatteten Geräte eine Vielzahl an Daten, wie Bilder und Videos, Tonaufnahmen, Kommunikationsströme, Positionsdaten und vieles mehr. Für Smartphone-Hersteller, Betriebssystem-Entwickler, App-Anbieter und Werbetreibende sind diese Daten sowohl zur Verbesserung ihrer Produkte als auch für die Optimierung von Produktwerbung attraktiv. Mit Hilfe von immer mächtigeren Werkzeugen und Algorithmen lassen sich aus diesen Daten erstaunlich ausführliche

[1] Universität Siegen, Zentrum für ökonomische Bildung in Siegen (ZöBiS), Kohlbettstraße 17, 57072 Siegen, froitzheim@zoebis.de
[2] Universität Siegen, Zentrum für ökonomische Bildung in Siegen (ZöBiS), Kohlbettstraße 17, 57072 Siegen, schuhen@zoebis.de
[3] Universität Siegen, Zentrum für ökonomische Bildung in Siegen (ZöBiS), Kohlbettstraße 17, 57072 Siegen, stentenbach@zoebis.de

doi:10.18420/inf2019_79

Personenprofile bilden, in denen private Informationen wie Beziehungen, Interessen und Hobbies, Gesundheitszustand, Bewegungsprofile etc. enthalten sein können. Obwohl den Verarbeitern der Daten damit tiefe Einblicke in die Privatsphäre der Verbraucher möglich sind, ist die Profilbildung aus den personenbezogenen Daten für die Verbraucher nicht transparent, da sie keinen direkten Einblick in die Verarbeitung ihrer Daten haben. Auch die Datenschutzgrundverordnung (DSGVO) hat die Datenverarbeitung für nicht informatisch gebildete Menschen nicht wirklich verbessert.

Der Schutz vor Missbrauch dieser sensiblen Daten soll in Deutschland durch das Zusammenwirken verschiedener Gesetze gewährleistet werden. Begründet wird dieser Schutz durch das im Volkszählungsurteil [BVerfG83] als Grundrecht anerkannte Recht auf informationelle Selbstbestimmung [Ho15], also dem Recht jedes Einzelnen über die Preisgabe und Verwendung seiner personenbezogenen Daten selbst zu bestimmen. In der Praxis wird dies meist durch eine Einwilligung des Verbrauchers umgesetzt.

Solche Einwilligungen werden alltäglich von Verbrauchern erteilt, wenn sie Apps auf ihren Smartphones, Dienstleistungen im Internet und vieles mehr nutzen. Ziel dieses Beitrages ist es mit Hilfe einer interaktiven simulationsgestützten Lerneinheit die oft überfordernden und verwirrenden Einwilligungsprozesse für Schülerinnen und Schüler verständlicher zu machen, und sie damit zu befähigen, souverän über ihre Daten verfügen zu können bzw. sie zu sensibilisieren für die möglichen Gefahren, die durch die digitale Welt entstehen.

2 Darstellung der Lerneinheit

2.1 Rahmenbedingungen

Im Rahmen des Forschungsvorhaben wurde eine Lerneinheit für den Einsatz in der Mittelstufe konzipiert. Vorrangig ist sie zur Einbettung in Unterrichtsreihen zum Thema Verbraucherschutz im Fach Sozialwissenschaften vorgesehen, wodurch der Schwerpunkt auch im Verbraucherschutz und der reflektierten Nutzung informatischer Systeme liegt.

Damit die Lerner ohne Gefahr, dass echte Daten abgegriffen oder tatsächlich Apps gekauft werden, die Inhalte erlernen können, wurde eine Simulation eines Smartphones entwickelt. Es werden nur grundlegende Fähigkeiten in der Bedienung von Smartphones vorausgesetzt, um die Simulationen bedienen zu können. Diese Fähigkeiten sollten als gegeben zu erwarten sein, da nahezu jeder Schüler dieser Altersgruppe ein eigenes Smartphone besitzt [FPR17] und den Umgang damit gewohnt sein müsste. Das in der Simulation simulierte Smartphone ist nicht als Abbild eines real existierenden Systems gedacht, sondern eine Umsetzung der zentralen Konzepte der verschiedenen real existierenden Systeme.

Jeder Schüler sollte Zugang zu einem eigenen Computer oder Tablet haben, um die Lerneinheit eigenständig durchführen zu können. Jedes Gerät sollte dabei mit dem Internet verbunden sein, damit es mit dem (Datenbank-)Server kommunizieren kann. Die Simulation ist in das elektronische Schulbuch ECON EBook integriert worden, um die Simulation in einer den Schülern bekannten Umgebung bereitzustellen. [SF15, FS15]

3 Entwicklung der Simulation

3.1 Anforderungen und Inhalt

Ziel der Simulation ist es, die Schüler mit realitätsnahen Handlungssituationen zu konfrontieren und sie Entscheidungen treffen zu lassen, die sie im weiteren Verlauf der Lerneinheit reflektieren. Dafür ist es notwendig, die von den Schülern getroffenen Entscheidungen zu dokumentieren und zur Visualisierung aufzubereiten. Es muss außerdem möglich sein, Schülergruppen zu bilden und die Entscheidungen der gesamten Gruppe darzustellen, um die Schüler zur Reflexion anzuregen und gegebenenfalls auch im Plenum die Unterschiede im Verhalten der Schüler aufzuzeigen.

Die Simulation sollte sich möglichst realistisch bedienen lassen und auch optisch klar als Simulation eines Mobiltelefons erkennbar sein. Dazu gehört auch, dass Dialoge und Anzeigeelemente möglichst nicht abstrakt, sondern so nah wie möglich an ihren originalen Vorbildern orientiert sind. Da sich die verschiedenen Mobiltelefone jedoch in Größe und Betriebssystem unterscheiden, ist es nicht möglich ein exaktes Abbild aller Mobiltelefone der Schüler zu erstellen. Stattdessen soll die Simulation die Konzepte der am meisten verbreiteten Betriebssysteme aufgreifen und so kombinieren, dass die Parallelen zu diesen originalen Betriebssystemen immer erkennbar sind.

Die Gestaltung der Simulation soll auch ihre Verwendbarkeit für andere Unterrichtseinheiten und Forschungsvorhaben ermöglichen, daher muss sie möglichst unkompliziert anzupassen und zu erweitern sein.

Die weiteren Anforderungen an die Simulation lassen sich in inhaltliche und technische Anforderungen unterteilen. Die inhaltlichen Anforderungen ergeben sich aus den didaktischen Entscheidungen und den Lernzielen, die technischen Anforderungen wiederum aus den didaktischen und inhaltlichen Anforderungen.

3.2 Inhaltliche Anforderungen

Die Simulation soll unter anderem zur Reflektion des Umgangs mit WLAN-Netzwerken und Berechtigungen anregen. Es ist notwendig eine realistische Darstellung des Verhaltens eines WLAN-Netzwerkes zu simulieren. Die Schüler müssen die Möglichkeit haben, ihr virtuelles Mobiltelefon mit verschiedenen gesicherten oder nicht gesicherten WLAN-Netzwerken zu verbinden. Zu einer realistischen Darstellung von WLAN auf

Mobiltelefonen gehört die Möglichkeit, das WLAN ein- und ausschalten zu können, Verbindungen aufzubauen und zu trennen, sowie das Einstellen der automatischen Verbindung mit gespeicherten Netzwerken. Die dafür notwendigen Menüs sollen ihren realen Vorbildern möglichst nahekommen und müssen daher auch noch weitere Einstellungsmöglichkeiten, wie das Ein- und Ausschalten von Bluetooth und Flugzeugmodus haben, da auch auf realen Mobiltelefonen viele Einstellungen gruppiert dargestellt werden.

Nach dem gleichen Prinzip soll die Simulation auch auf dem Startbildschirm und in anderen Menüs Möglichkeiten zur Interaktion bieten. Dadurch soll einerseits die bereits erwähnte realistische Darstellung gewahrt werden und andererseits die Möglichkeit zur Exploration gegeben werden.

Wenn die Schüler ihr WLAN aktiviert und eine Verbindung mit einem Netzwerk hergestellt haben, sollen sie den App Store des Telefons aufrufen können und eine App herunterladen können. Das Herunterladen der App über das vorher vom Schüler ausgewählte Netzwerk verknüpft die Frage der Datensicherheit im WLAN mit der des Datenschutzes in Apps. Bevor die App heruntergeladen und installiert werden kann, muss es möglich sein alle Berechtigungen, die die App anfordern kann einzusehen. Es muss erkennbar sein, dass die Berechtigungen zu diesem Zeitpunkt noch nicht gewährt werden, sondern lediglich für den Gebrauch der App später angefordert werden können.

Nach der Installation bittet die App beim ersten Öffnen um mehrere Berechtigungen, von denen mindestens eine offensichtlich für den Betrieb der App benötigt wird und eine auch abgelehnt werden kann, ohne die Funktionalität der App zu beeinträchtigen.

3.3 Technische Anforderungen

Um der heterogenen Ausstattung von Schulen gerecht zu werden, muss die Simulation flexibel auf unterschiedlichen Geräten und Betriebssystemen einsetzbar sein. Die Eingaben der verschiedenen Schüler müssen persistent gespeichert, kombiniert und verarbeitet werden können. Dafür ist es notwendig Identifikationsmerkmale zu definieren, mit deren Hilfe Schülerantworten und Entscheidungen innerhalb der Simulationen den jeweiligen Schülern zugeordnet werden können. Die Darstellung der unterschiedlichen Entscheidungen innerhalb einer Gruppe ist nur möglich, wenn die Gruppenzugehörigkeit für jeden Schüler eindeutig zugeordnet werden kann, daher muss ein entsprechendes Attribut zu jedem Schüler abgespeichert werden können.

3.4 Rahmenbedingungen und verwendete Technologien

Die Simulation wurde als Anwendung für den Webbrowser entwickelt. Der Webbrowser ermöglicht es HTML-basierte Internetseiten auf den Geräten der verschiedenen Betriebssysteme gleich oder in an die jeweilige Bildschirmgröße angepasster Version dar-

zustellen. Diese Tatsache legt die Implementierung der Simulation als browserbasierte Webanwendung (Web-App) nahe.

Web-Apps basieren auf dem Client-Server-Modell. Die Anwendungen müssen für den Gebrauch nicht auf den lokalen Rechnern der Benutzer installiert werden, sondern befinden sich auf einem Webserver. Der Webbrowser (der Client im Client-Server-Modell) ermöglicht es auf die Web-App zuzugreifen und sie auf dem jeweiligen Gerät anzuzeigen. Die Datenverarbeitung kann sowohl auf dem Server als auch auf dem Client stattfinden. In die gängigen Webbrowser ist dafür eine JavaScript-Umgebung integriert, die Daten selbst verarbeiten und mit einem Webserver austauschen kann. Auf dem Webserver wiederum lassen sich verschiedene Skriptsprachen wie PHP, Perl, Python oder Ruby verwenden. Da die Größe des Projektes relativ gering ist, spielt bei der Wahl der Skriptsprache die Komplexität ihrer Konfiguration und Nutzung die entscheidende Rolle, weshalb für die Simulation PHP verwendet wird.

Ein weiterer Vorteil der Implementierung als Web-App ist die Möglichkeit zur zentralen Datenspeicherung auf dem Webserver. Vergleiche innerhalb von Lerngruppen sind eine der genannten Anforderungen, die durch eine zentrale Datenbank gelöst wird. Als Datenbanktechnologie wird MySQL verwendet, da es sich mit geringem Aufwand auf einem Webserver verwalten lässt und eine leicht verwendbare Schnittstelle mit PHP besitzt.

Auf Seite des Clients wird die Kompatibilität mit den verschiedenen Webbrowsern durch den Verzicht auf Browser-Plugins (wie z.B. Flash) gewährleistet. Hier werden lediglich HTML und CSS für die Anzeige und JavaScript für die Interaktivität verwendet. Unterstütz wird der Client dabei durch die freie, plattformunabhängige JavaScript-Bibliothek jQuery Mobile. JQuery Mobile ermöglicht es Benutzeroberflächen zu schaffen, die optisch an das Design von Betriebssystemen auf Mobiltelefonen angelehnt sind und sich sowohl mit Maus und Tastatur als auch mit Berührungen und Gesten auf Touchscreens steuern lassen.

Die Visualisierung der Entscheidungen innerhalb von Gruppen wird mit der ebenfalls plattformunabhängigen JavaScript-Bibliothek Highcharts ermöglicht.

Alle eingebundenen grafischen Elemente, wie Icons und Symbole, die nicht selbst erstellt wurden, stammen aus verschiedenen Quellen und sind mit bedingungslosen Lizenzen (CC0) versehen und damit frei verwendbar.

3.5 Architektur

Das System ist aufgegliedert in drei, auf Client und Server verteilte Schichten. Die Datenhaltung befindet sich vollständig auf der Serverseite. Die Präsentation findet ausschließlich beim Client statt. Die Logikschicht als Bindeglied zwischen Präsentation und Datenhaltung verteilt sich über Client und Server. Der Client verarbeitet die Eingaben der Benutzer und nutzt die Serverseitige API, um datennahe Berechnungen durchführen zu lassen und Daten zur Speicherung zu übergeben, oder aus der Datenbank zu empfangen.

Abb. 1: Startseite der Simulation

4 Darstellung der entwickelten Simulation

Die entwickelte Simulation enthält verschiedene Ansichten, durch die sich der Benutzer navigieren kann. Zu diesen Ansichten gehören die Startseite, die einzelnen Seiten der Einstellungen des App Stores und die simulierten Apps „Firefox“ als Browser und die „Uhr“. Die Navigation findet durch Klicken oder Berühren von Icons und Schaltflächen innerhalb der Simulation statt. Je nach Navigationsrichtung durchgeführte Animationen beim Seitenübergang ermöglichen es zu erkennen, ob eine tiefere oder höhere Ebene einer App geöffnet wurde.

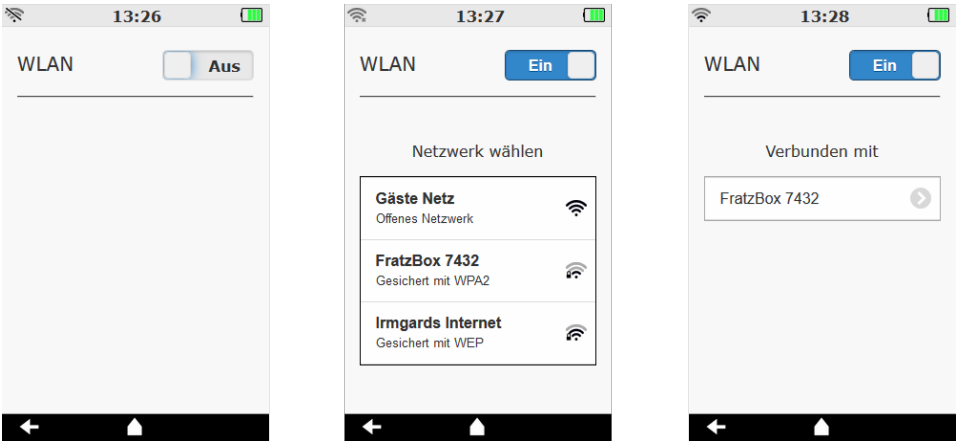


Abb 2: Die WLAN Einstellungen bei ausgeschaltetem, nicht verbundenem und verbundenem WLAN

4.1 Status- und die Navigationsleiste

Im oberen Bereich der Simulation wird eine Status-Leiste mit Symbolen für den Netzwerkstatus und den Akkustand sowie eine Uhr angezeigt. Diese Status-Leiste ist in allen Ansichten vorhanden. Die angezeigte Uhrzeit entspricht der vom Endgerät des Benutzers vorgegebenen Uhrzeit und läuft in Echtzeit weiter. Das Symbol für den Netzwerkstatus entspricht den bekannten WLAN-Symbolen und kann sich durch Benutzereingaben in den Einstellungen verändern. Beim ersten Start der Simulation ist das Symbol ausgegraut und mit einem kleinen Kreuz versehen, um zu zeigen, dass das WLAN nicht verbunden ist. Schaltet der Benutzer das WLAN des simulierten Mobiltelefons aus, wird das Symbol ausgegraut und durchgestrichen angezeigt.

Die Simulation beginnt mit der Anzeige eines Startbildschirms, von dem aus zu den Apps „Firefox“, „Uhr“, dem App Store sowie zu den Einstellungen des simulierten Mobiltelefons navigiert werden kann. Die Navigation findet durch das Klicken bzw. Berühren der beschrifteten Icons statt. Der Hintergrund des Startbildschirms ist mit einem Hintergrundbild geschmückt.

Die erste Ansicht nach dem Öffnen der Einstellungen beinhaltet einen Schalter zum Ein- und Ausschalten des Flugmodus und Navigations-Elemente zu den Einstellungen für WLAN, Bluetooth und den Einstellungen der Apps „Uhr“ und „Firefox“. Der Schalter für den Flugmodus ist beim Start ausgeschaltet. Wird er aktiviert, werden WLAN und Bluetooth deaktiviert. Wird der Flugmodus wieder deaktiviert, gelangen WLAN und Bluetooth wieder in den Zustand zurück, in dem sie sich vor Aktivieren des Flugmodus befanden. Eine gegebenenfalls aktive WLAN-Verbindung wird allerdings nur reakti-

viert, wenn die automatische Verbindung für dieses Netzwerk ausgewählt wurde. Wie bei einem echten Mobiltelefon lässt sich auch in der Simulation WLAN und Bluetooth aktivieren, während der Flugmodus aktiv ist.

Durch das Auswählen der mit „Bluetooth“ beschrifteten Schaltfläche wird eine neue Ansicht geöffnet, auf der sich der Ein- und Ausschalter für das Bluetooth befindet. Wird das Bluetooth eingeschaltet, wird im Bereich darunter der Schriftzug „Ihr Gerät ist jetzt sichtbar als 'Mein_Phone'.“ angezeigt.

Die „Firefox“-Einstellungen erlauben es sowohl eine Suchmaschine auszuwählen als auch auszuwählen, ob Webseiten beim Öffnen eine Do-Not-Track-Nachricht übermitteln dürfen oder nicht. Darüber wird die dazugehörige Erklärung „Websites immer mitteilen, meine Nutzeraktivitäten nicht zu verfolgen“ angezeigt. Als Suchmaschinen stehen „Google“, „Bing“, „Yahoo“ und „DuckDuckGo“ zur Auswahl.

Die WLAN-Einstellungen stellen das Kernstück der Einstellungen dar. Ihre Startseite (vgl. Abb. 2) zeigt bei ausgeschaltetem WLAN nur den Schalter zum Ein- oder Ausschalten des WLAN an. Sobald das WLAN eingeschaltet wird, erscheint eine Liste der Verfügbaren Netzwerke, zu denen jeweils die Verschlüsselung und die Empfangsstärke angezeigt werden. Zur Auswahl stehen das „Gäste Netz“, ein offenes Netzwerk mit vollem Empfang, das Netzwerk „FratzBox7432“, gesichert mit WPA2, aber nur einem von drei Empfangsbalken und „Irmgards Internet“, ein mit WEP gesichertes Netzwerk mit zwei von drei Empfangsbalken.

Durch Auswählen eines Netzwerks gelangt der Benutzer in die nächste Ansicht, in der er sich mit dem jeweiligen WLAN verbinden kann. Dazu befinden sich in dieser Ansicht jeweils der Name des Netzes, ein Auswahlkasten zum automatischen Verbinden und ein Button mit der Aufschrift „Verbinden“. Die beiden gesicherten Netzwerke haben außerdem noch ein Feld für die Eingabe des Passworts. Bleibt dieses Feld leer, oder ist dort ein falsches Passwort eingegeben, ist eine Verbindung mit dem jeweiligen Netzwerk nicht möglich und es wird in roter Farbe der Hinweis „Passwort inkorrekt“ über dem Feld angezeigt. Wenn das ggf. abgefragte Passwort korrekt ist, wird durch das Betätigen des „Verbinden“-Buttons wieder zur vorherigen Ansicht navigiert, auf der daraufhin nur noch das Netzwerk ausgewählt werden kann, mit dem der Benutzer gerade verbunden wurde. Durch erneutes Auswählen des Netzwerks gelangt der Benutzer wieder in die Ansicht, in der er sich vorher verbunden hat. Dort befinden sich nun nur noch der Auswahlkasten und ein Button mit der Aufschrift „Trennen“. Mit diesem Button wird die Verbindung wieder getrennt und alle WLAN-Netzwerke stehen erneut zur Auswahl. Wird das WLAN während einer Verbindung zu einem Netzwerk besteht deaktiviert, wird die Verbindung getrennt. Nach erneutem Aktivieren des WLAN wird die alte Verbindung wiederhergestellt, falls der Auswahlkasten für das automatische Verbinden nicht abgewählt wurde.

4.2 Die „Firefox“-App

Die „Firefox“ App simuliert einen mobilen Webbrowser. Die normalerweise für einen Browser essentielle Adressleiste ist ausgeblendet, da deren sinnvolle Implementierung unverhältnismäßig komplex für die Anwendungsszenarien der Simulation wäre. Um dennoch Internet-Funktionalität zu simulieren, besitzt sie eine Suchmaschine als Startseite. Standardmäßig ist hier „Google“ als Suchmaschine eingestellt, in den Einstellungen lassen sich aber auch „Bing“, „Yahoo“ und „DuckDuckGo“ auswählen. Die Suchmaschine stellt ein Eingabefeld und einen Button mit der Beschriftung „Suchen“ zur Verfügung. Sucht der Benutzer nach etwas, wird im Hintergrund überprüft ob die Eingabe die Buchstaben- Kombination zum Beispiel „laufta“ enthält. Ist dies der Fall, wird als Suchergebnis ein Treffer mit dem Inhalt „Lauftastic - Jetzt im App Store verfügbar“ angezeigt, der als Link in den App Store fungiert. Dieser Mechanismus dient als Hilfestellung für Schüler, die versuchen die App „Lauftastic“ im Browser, statt im App Store herunterzuladen. Ist im Suchbegriff die Buchstaben-Kombination nicht in der Eingabe vorhanden wird angezeigt, dass keine Ergebnisse für den gewählten Suchbegriff gefunden wurden.

Wie ein realer Webbrowser funktioniert auch die simulierte App nur bei bestehender Internetverbindung. Ist kein WLAN verbunden, zeigt der Browser die Fehlermeldung „Server nicht gefunden“ an.

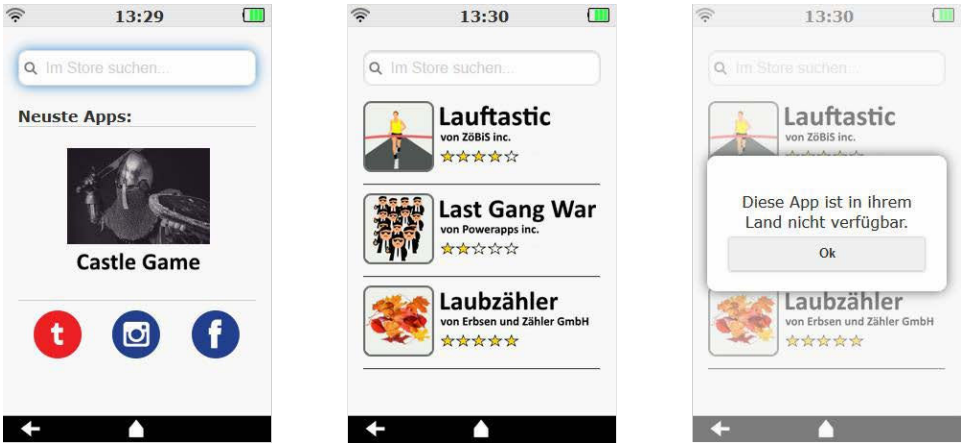


Abb. 3: Startseite, Suche und nicht verfügbare App im simulierten App Store

4.3 Der App Store

Der App Store bietet die Möglichkeit neue Apps zu entdecken und auf dem simulierten Mobiltelefon zu installieren. Auf der Startseite (Abb. 3 links) befindet sich ein Eingabefeld zum Suchen von Apps, eine rotierende Ansicht der neuesten Apps und App-Icons. Alle dargestellten Apps mit Ausnahme der „Lauftastic“ App können nicht geöffnet wer-

den, an Stelle der jeweiligen Shop-Seite wird für diese Apps nur ein Hinweisfenster mit der Meldung, dass die App nicht verfügbar ist, angezeigt, wenn versucht wird, sie zu öffnen (Abb. 3 rechts). Das Suchfeld zeigt Treffer-Vorschläge, sobald etwas darin eingegeben wird. Hier sollen die Schüler nach der „Lauftastic“ App suchen können. Durch das Auswählen des Treffer-Vorschlags wird eine Seite mit Suchergebnissen geöffnet, auf der unter „Lauftastic“ auch die Apps „Last Gang War“ und „Laubzähler“, jeweils mit Hersteller und Bewertungen auf einer Skala von einem bis fünf Sternen, angezeigt werden (Abb. 3 mittig). Das Auswählen von „Last Gang War“ und „Laufzähler“ führt zur Anzeige des Hinweises, dass die App nicht verfügbar ist.

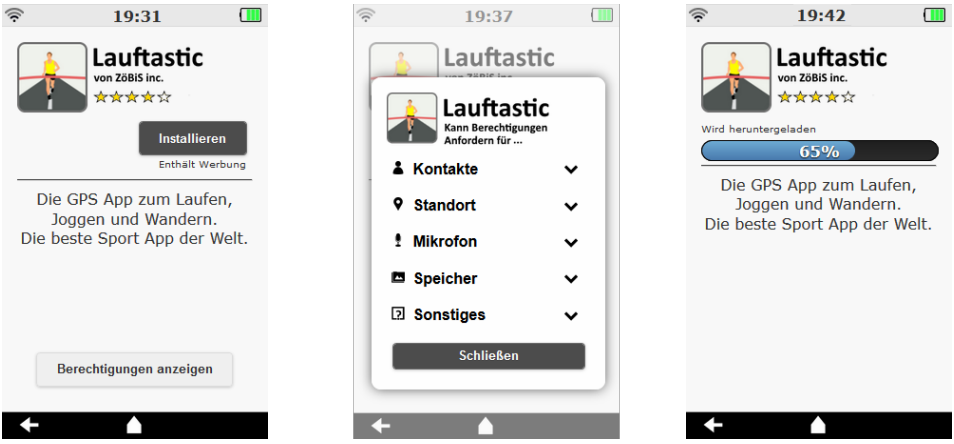


Abb. 4: Shop-Seite der „Lauftastic“ App, mögliche angeforderte Berechtigungen und Installation der App

Wenn „Lauftastic“ ausgewählt wird, öffnet sich die zugehörige Shop-Seite der App, auf der eine Kurzbeschreibung, die Bewertungen, der Hinweis „Enthält Werbung“ und ein Button mit der Aufschrift „Installieren“ angezeigt werden (Abb. 4 links). Im unteren Bereich befindet sich ein Button zum Anzeigen der anforderbaren Berechtigungen der App. Mit einem Klick auf diesen Button öffnet sich ein Hinweisfenster, in dem die Berechtigungen aufgelistet sind (Abb. 4 mittig). Die Berechtigungen sind in die Gruppen „Kontakte“, „Standort“, „Mikrofon“, „Speicher“ und „Sonstiges“ unterteilt. Durch einen Klick auf eine Gruppe öffnet oder schließt sich darunter eine Auflistung der einzelnen Berechtigungen in dieser Kategorie.

Die Installation der App „Lauftastic“ lässt sich durch Auswählen der „Installieren“-Schaltfläche und anschließendes Bestätigen des Installationswunsches starten. Der Installationsprozess wird durch einen Fortschrittsbalken simuliert (Abb. 4 rechts).

Im weiteren Verlauf der Unterrichtsreihe können die Schüler die App Lauftastic öffnen und weitere InApp-Käufe tätigen und die Berechtigungen verändern.

5 Visualisierung und Reflexion des Schülerverhaltens

Die Schüler erhalten über das elektronische Schulbuch Arbeitsaufträge, in denen sie zum Beispiel eine geeignete WLAN-Verbindung herstellen sollen. Die Aufgabenstellungen sind bewusst offen gestaltet, damit die Schüler abwägen müssen, welche Verbindung geeignet ist, um zum Beispiel sicher Daten zu transferieren. Die während der Bearbeitung durchgeführten Arbeitsschritte und Entscheidungen werden in einer Datenbank gesammelt. Die in der Datenbank gesammelten Informationen zum Verhalten der Schüler innerhalb einer Lerngruppe können über die Server-API abgefragt und mit Hilfe der JavaScript-Bibliothek Highcharts als beschriftete Tortendiagramme angezeigt werden. Sie können als HTML-Elemente auf verschiedenen Plattformen eingebettet werden, oder über eine auf dem Server verfügbare Internetseite abgerufen werden. Diese Internetseite ermöglicht auch die Diskussion der Ergebnisse im Unterrichtsgespräch zwischen Lehrer und Schülern. Im Rahmen der Reflexion wird den Schülern die Gefahren bewusst und auch ein Bezug zur Realität hergestellt.

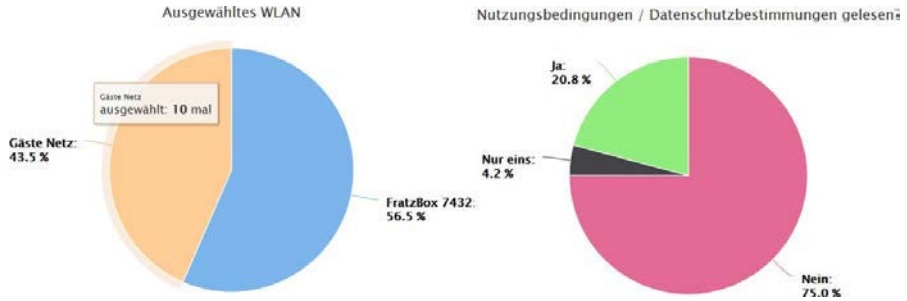


Abb. 5: Dynamisch generierte Tortendiagramme zur Visualisierung des Schülerverhaltens innerhalb der Simulation

6 Evaluierung

Die Simulation wurde in mehreren Lerngruppen der Jahrgangsstufe 10 in Nordrhein-Westfalen mit insgesamt 76 Lernenden eingesetzt. Vor und nach dem Einsatz der Simulation wurde jeweils ein Fragebogen zur Selbsteinschätzung und zu den Kompetenzen von den Lernenden bearbeitet. Die Selbsteinschätzungen der Schüler konnten innerhalb der Simulation unter Beweis gestellt werden. Hierdurch wird auch die Differenz zwischen den vorhandenen Kompetenzen und der konkreten Handlung deutlich.

Es gelang allen Schülern, wie von ihnen selbst erwartet, eine Verbindung zu einem der Netzwerke herzustellen. Nur 54 Prozent der Schüler wählten die Verbindung mit dem sicheren Netzwerk. Die anderen Schüler wählten das offene, nicht gesicherte Netzwerk. Als Grund gaben viele Schüler in der Diskussion an, dass die Bequemlichkeit der Hauptgrund für die Entscheidung war.

Bis auf einen Schüler waren alle Schüler vor und nach der Unterrichtsstunde überzeugt davon, eine App auf ihrem Smartphone installieren zu können, was die meisten auch innerhalb der Simulation demonstrieren konnten. Lediglich vier Schülern gelang es nicht im entsprechenden Simulationsabschnitt die App „Lauftastic“ zu installieren. Alle vier hatten sich jedoch dazu entschieden die Liste der möglichen Berechtigungsanfragen abzulehnen und daher möglicherweise ganz bewusst auf das Herunterladen der App verzichtet. Die anderen Schüler haben den allgemeinen Geschäftsbedingungen der App und den geforderten Berechtigungen unreflektiert zugestimmt.

Die Fähigkeit Datenschutzeinstellungen auf einem Smartphone einzurichten ist im Gesamten relativ komplex. Unter anderem können innerhalb des Browsers und des Betriebssystems darüberhinausgehende Einstellungen vorgenommen werden. Ein möglicherweise positiver Hinweis ist die Tatsache, dass ein Großteil der Schüler das Recht Mitteilungen zu senden, in dem Bewusstsein damit nervige Benachrichtigungen zu verhindern, abgelehnt hat. Die Entscheidung den Standortzugriff zu erlauben oder zu verbieten, kann jedoch nicht als Hinweis gesehen werden, da es hier keine objektiv richtige Entscheidung gibt. Allerdings wurde aus der Befragung deutlich, dass nicht über die Folgen dieser Entscheidung im Vorfeld nachgedacht wurde.

Zusätzlich zu den Handlungen innerhalb der Simulation lässt sich mit Hilfe der Wissensfragen der Wissenszuwachs nach der Nutzung der Unterrichtseinheit messen. In allen Wissensfragen zeigt sich eine deutliche Verbesserung des Wissenstandes und eine deutlich geringere Unsicherheit bei den Schülern.

Die Frage, ob in einem offenen WLAN-Netzwerk alle Daten verschlüsselt übertragen werden, beantworteten vorher 57% der Schüler korrekt. 35% gaben an nicht sicher zu sein und 8% beantworteten die Frage falsch. Nach der Unterrichtsstunde waren lediglich 3% unsicher, 74% beantworteten die Frage korrekt.

Ob innerhalb eines passwortgeschützten WLAN-Netzwerkes alle Daten verschlüsselt übertragen werden, beantworteten vor der Unterrichtsstunde 54% der Schüler korrekt, 14% beantworteten die Frage falsch und 32% gaben an unsicher zu sein. Nach der Unterrichtsstunde beantworteten 89% der Schüler die Frage richtig, während jeweils 6% unsicher waren oder die falsche Antwort auswählten.

Die Frage, ob unverschlüsselt übertragene Daten von allen in Reichweite liegenden WLAN- Geräten mitgelesen und gespeichert werden können, beantworteten im Pretest lediglich 35% der befragten Schüler richtig. 30% der Schüler hielten dies fälschlicherweise für nicht möglich und 35% der Schüler waren sich unsicher. Im Posttest antworteten 83% der Schüler korrekt, 6% lagen mit ihrer Antwort nicht richtig und 11% waren noch immer unsicher. Die Frage hätte jedoch auch präziser formuliert werden müssen, da es unwahrscheinlich, aber nicht unmöglich wäre, sie so zu interpretieren, dass auch Datenflüsse über andere Medien, wie Kupfer- oder Glasfaserkabel gemeint sein könnten.

Ob die mit Apps gesammelten Daten unter keinen Umständen an Dritte, wie andere Unternehmen oder Werbepartner weitergegeben werden dürfen, konnten vor der Unter-

richtsstunde 32% der Schüler richtig beantworten. 41% beantworteten die Frage falsch und 27% gaben an unsicher zu sein. Nach der Unterrichtsstunde wussten 49% der Schüler die richtige Antwort. Der Anteil der falschen Antworten sank auf 40% und der Anteil der unsicheren Schüler lag bei 11%.

43% der Schüler gaben richtigerweise vor der Unterrichtsstunde an, dass Apps auch dann nicht grundsätzlich unbedenklich sein müssen, wenn sie im offiziellen App Store angeboten werden. 27% beantworteten diese Frage falsch und 30% konnten keine sichere Antwort geben. Nach der Unterrichtsstunde beantworteten 49% die Frage korrekt, 31% wählten die falsche Antwort und 20% waren noch immer unsicher.

In der letzten Frage wurden die Schüler gefragt, ob auf dem eigenen Mobiltelefon installierte Apps nur für den Nutzer selbst eine Gefahr darstellen können. Im Pretest antworteten 30% der Schüler mit der korrekten Antwort, 46% von ihnen antworteten mit der falschen und 27% gaben an unsicher zu sein. Im Posttest antworteten 46% korrekt, 40% inkorrekt und nur 14% waren noch immer unsicher.

Die ersten drei der oben vorgestellten Fragen beziehen sich auf das Thema Datensicherheit in WLAN-Netzwerken. Sie wurden im Zuge des Unterrichtsgesprächs nach Durchführung des ersten Simulationsabschnittes thematisiert. In allen drei Fragen sind erhebliche Steigerungen im Wissensstand der Schüler zu erkennen. Nach der Unterrichtsstunde sind nur noch 10% der Schüler davon überzeugt, dass in offenen WLAN-Netzwerken alle Daten verschlüsselt übertragen werden.

7 Fazit

Im vorliegenden Aufsatz wurde eine Lerneinheit zum Thema „Datenschutz als Verbraucherschutz“ entwickelt und ihre Kernelemente im Schulunterricht erprobt. Ziel der Lerneinheit ist es, die Schüler in ihrer Rolle als Verbraucher digitaler Dienstleistungen zu stärken, indem ihre Fähigkeit zum souveränen Umgang mit den eigenen Daten geschult wird. Dazu wird exemplarisch die Installation und Verwendung von Apps auf Smartphones betrachtet und in der handelnden Auseinandersetzung mit der Simulation eines Smartphones erforscht, wie Daten auf einem solchen Gerät gesammelt werden, warum an diesen Daten ein kommerzielles Interesse bestehen kann und wie das Sammeln der Daten eingeschränkt werden kann. Als weiterer wichtiger Aspekt zum Schutz der eigenen Daten wird auch das Thema Datensicherheit in WLAN-Netzwerken thematisiert.

Die Lerneinheit ist an der Lebensrealität der Schüler ausgerichtet, um den Schülern die Übertragung des Gelernten auf reale Situationen zu vereinfachen. Durch die Integration der Simulationen enthält sie einen weiteren Motivationsanreiz, der in der Erprobung der Lerneinheit von den Schülern bestätigt wurde.

Um den möglichst flexiblen Einsatz der Lerneinheit zu ermöglichen, lassen sich die Simulationen ohne großen Aufwand in verschiedene Lernplattformen und andere Werkzeuge einbetten. Zwischen den einzelnen Abschnitten befindet sich frei gestaltbarer Raum, der es ermöglicht weitere Themen und Schwerpunkte zu integrieren und die Lerneinheit an die jeweilige Situation anzupassen.

Die Erprobung zeigte, dass die Zielgruppe der Lerneinheit mit den Simulationen gut umgehen und Lernerfolge erzielen konnte. Insbesondere stellt die Reflexion der eigenen Entscheidungen in der Simulation einen wesentlichen Beitrag dar, um die Mündigkeit der Verbraucher in der digitalen Welt in der Breite zu fördern.

Literaturverzeichnis

[BVerfG83] Urteil des Ersten Senats vom 15. Dezember 1983, 1 BvR 209/83 u. a. – Volkszählung –, BVerfGE 65, 1.

[FPR17] Feierabend, S.; Plankenhorn, T.; Rathgeb, T. (2017). JIM 2017: Jugend, Information, (Multi-)Media. Stuttgart Medienpädagogischer Forschungsverbund Südwest (mpfs).

[FS15] Froitzheim, M./ Schuhen, M. (2015): Das ECON EBook als interaktives und multimediales elektronisches Schulbuch für den Ökonomieunterricht. In: Pongratz, H./ Keil, R. (Hrgs.): DeLFI 2015 - Die 13. E-Learning Fachtagung Informatik der Gesellschaft für Informatik. Bonn: Köllen Druck+Verlag GmbH. S. 253-264.

[Ho15] Hornung, G. (2015). Grundrechtsinnovationen. Tübingen. Mohr Siebeck

[SF15] Schuhen, M.; Froitzheim, M. (2015): Konzeption des ECON EBooks mit dem Fokus „Gute Aufgaben“. In: Schuhen, M.; Froitzheim, M. (Hrsg.): Das Elektronische Schulbuch. Fachdidaktische Anforderungen und Ideen treffen auf Lösungsvorschläge der Informatik. Münster: LIT Verlag. S. 139-156.

Entwicklung eines theoretischen Rahmenwerks zur Erfassung von Medienkompetenz innerhalb von E-Learning-Systemen in der beruflichen Bildung

Konzeption, Evaluation und Ausblick in der Domäne des Stuckateur-Handwerks

Kim Petry[1], Tobias Greff[2], Dirk Werth[3]

Abstract: Diese Arbeit beschäftigt sich mit der Konzeption eines theoretischen Rahmenwerks, mittels dessen die Erfassung von Medienkompetenz in E-Learning-Systemen zum Zweck der Medienkompetenzvermittlung ermöglicht wird. Die Abhandlung ist im Kontext des vom BMBF geförderten Projekts D-MasterGuide entstanden. Ziel des Projekts ist es, Medienkompetenz durch die aktuelle Meisterausbildung zukünftig in die Stuckateurbetriebe zu bringen. Zunächst wurden eine strukturierte Literaturrecherche und Experteninterviews durchgeführt. Die daraus gewonnenen wissenschaftlichen Grundlagen zur Medienkompetenzvermittlung und -entwicklung wurden in einem aggregierten, generischen Medienkompetenzmodell vereint. Dieses wurde anschließend in die Domäne des Stuckateurhandwerks übertragen, im Rahmen eines Workshops evaluiert und unter Berücksichtigung der Evaluationsergebnisse überarbeite. Ergebnis der Arbeit ist ein umfassendes und praxisevaluiertes Medienkompetenzmodell mit 8 Dimensionen und 23 Kompetenzen. Darauf aufbauend wird analysiert und gezeigt, wie die Erfassung und Vermittlung von Medienkompetenz innerhalb eines E-Learning-Systems umgesetzt werden kann.

Keywords: Medienkompetenzmodell; handlungsorientierte Medienkompetenzvermittlung; Medienkompetenzerfassung; E-Learning; Handwerk

1 Einleitung

Die immer weiter voranschreitende Digitalisierung des Arbeitsalltags ist seit Jahren ein viel diskutiertes Thema, das stetig an Relevanz gewinnt. Mit Hilfe der Metasuchmaschine Google Scholar finden sich unter dem Suchbegriff „Digitalisierung der Arbeitswelt“ 2.700 wissenschaftliche Publikationen allein aus dem Jahr 2018. Dieser fortschreitende technische Wandel stellt neue Anforderungen an Arbeitnehmer und „erfordert Arbeitskräfte, die technologische Innovationen hervorbringen und nutzen können“ [EB15]. Der Umgang mit digitalen Technologien ist ein fester Bestandteil einer Vielzahl unterschiedlicher Berufe geworden [KJL15]. So verwundert es nicht, dass die Europäische Kommission die digitale

[1] AWS-Institut, Uni Campus Nord D 5 1, 66123 Saarbrücken, Deutschland kim.petry@aws-institut.de
[2] AWS-Institut, Uni Campus Nord D 5 1, 66123 Saarbrücken, Deutschland tobias.greff@aws-institut.de
[3] AWS-Institut, Uni Campus Nord D 5 1, 66123 Saarbrücken, Deutschland dirk.werth@aws-institut.de

doi:10.18420/inf2019_80

Medienkompetenz als Schlüsselkompetenz für lebensbegleitendes Lernen eingestuft hat [Eu06]. Der Begriff „Medienkompetenz“ umfasst Kompetenzen für ein souveränes Leben mit Medien. Diese beinhalten technische Fertigkeiten, Kommunikationskompetenzen und die Fähigkeit zur Reflexion und zum kritischen Umgang mit Medien [Sc09]. Damit bildet die Medienkompetenz die Grundlage digitaler Bildung.

Untersuchungen belegen allerdings, dass bei Schülerinnen und Schülern, Auszubildenden und Beschäftigten Defizite im Bereich der Medienkompetenz vorhanden sind [KJL15], [MM11], [BM16]. Trotzdem zeigte eine Studie aus dem Jahr 2016, dass 62% der Arbeitnehmer keine Weiterbildungen zur Erlangung digitaler Kompetenzen erhalten [Bi16]. Dieses Problem betrifft auch kleine und mittelständische Betriebe im Ausbauhandwerk, bei denen z. B. die Schnittstelle zwischen Baustellenmanagement und betrieblichem Backoffice viele Potenziale bietet. Durch fehlende Medienkompetenz und -akzeptanz bleiben diese häufig ungenutzt. Deshalb setzt das vom BMBF geförderte Projekt D-MasterGuide [EB19] bei der Meisterausbildung an, um Medienkompetenz durch die nächste Generation der Führungsebene in die Unternehmen zu bringen [BM19], [BM17]. Um zukünftige Meister auf die Digitalisierung in ihrem Beruf vorzubereiten, wird es dabei als essenziell erachtet, eine Form der Medienkompetenzentwicklung in das genutzte E-Learning-System – begleitend zu dessen fachlich-vermittelten Lerninhalten – zu integrieren.

Ziel dieser Arbeit ist die Konzeption eines theoretischen Rahmenwerks, mittels dessen die Erfassung von Medienkompetenz in E-Learning-Systemen zum Zweck der Medienkompetenzvermittlung ermöglicht wird. Bestehende wissenschaftliche Grundlagen zur Medienkompetenzvermittlung und -entwicklung sollen in einem aggregierten und praxisevaluierten Medienkompetenzmodell vereint werden. Das Modell soll anschließend in die spezielle Domäne des Stuckateurhandwerks übertragen und im Projekt D-MasterGuide eingesetzt werden. Insbesondere soll analysiert und gezeigt werden, wie Medienkompetenz während der handlungsorientierten Vermittlung von Lerninhalten innerhalb eines E-Learning-Systems erlangt werden kann. Ein langfristiger Einsatz im Stuckateurhandwerk und ähnlichen Domänen soll sichergestellt werden. Hierzu erfolgt in Kapitel 2 eine Literaturrecherche, in deren Rahmen etablierte Wissenschaftsdatenbanken strukturiert nach relevanter Literatur durchsucht werden. Als relevant gelten Arbeiten, die sich mit dem Stellenwert von Medienkompetenz für heutige Arbeitnehmer, Medienkompetenzmodellen, sowie Möglichkeiten der Messung, Erfassung und Vermittlung von Medienkompetenz beschäftigen. Ausgewählte Arbeiten werden systematisch erfasst und ihr Nutzen für diese Arbeit beschrieben. Anschließend werden in Kapitel 3 Experteninterviews zur Validierung und Ergänzung der Literaturrecherche durchgeführt. Kapitel 4 widmet sich – unter Einbeziehung der Ergebnisse der Literaturrecherche und der Experteninterviews – der Entwicklung eines domänenspezifischen Medienkompetenzmodells. Das Modell soll die Medienkompetenz, die im Stuckateurhandwerk benötigt wird, konkret und in ihrer Gesamtheit beschreiben und die Grundlage für die Medienkompetenzerfassung und -vermittlung innerhalb einer Lernplattform bilden. Darauf aufbauend werden in Kapitel 5 als Ausblick die Möglichkeiten

der Kompetenzerfassung innerhalb eines E-Learning-Systems erläutert. Zuletzt folgen in Kapitel 6 Fazit und Ausblick auf zukünftige Forschungen.

2 Literaturrecherche

Gemäß dem Vorgehen von Brocke et al. wird die vorhandene Literatur mittels fest definierter Suchstrings systematisch nach relevanten Beiträgen durchsucht [Br09]. Anschließend werden die Quellen initial priorisiert und ausgewählte Publikationen detailliert analysiert. Dabei wird eine Forward- und Backward-Suche durchgeführt. Relevante verwandte Arbeiten werden so identifiziert und wie von Webster und Watson vorgeschlagen kategorisiert [WW02]. Die definierten Suchstrings gliedern sich in drei Themenfelder:

- Medienkompetenz in der beruflichen Bildung:
  - S1: „Vermittlung von Medienkompetenz“ AND „berufliche* Bildung“
  - S2: „Medienkompetenzförderung“ AND „berufliche* Weiterbildung“
  - S3: „Medienkompetenzförderung“ AND „Berufsbildung“
  - S4: „Medienkompetenz“ AND „HandwerkÄND „berufliche* Weiterbildung“
- Medienkompetenz messen:
  - S5: „empirische Erfassung“ AND („MedienkompetenzÖR „Informationskompetenz“)
  - S6: „Messung von Medienkompetenz“ OR „Messung von Informationskompetenz“"
  - S7: „Testinstrumente“ AND („MedienkompetenzÖR „Informationskompetenz“)
- Medienkompetenzmodell:
  - S8: „Medienkompetenzmodell“ AND „Definition“
  - S9: „Kompetenzbündel“ AND („Medienkompetenz“ OR „Informationskompetenz“)

Die Recherche wurde mit der Metasuchmaschine Google Scholar durchgeführt, welche etablierte wissenschaftliche Datenbanken wie beispielsweise Springerlink, EBSCOhost, JSTOR, Elsevier, u. a. vereint. Aufgrund des schnellen technologischen Wandels und den damit einhergehenden sich ändernden Anforderungen an den Umgang mit Medien, wurde das Hauptaugenmerk auf aktuelle Publikationen (seit 2012) gelegt.

Die Ergebnisse (s. Tab. 1) wurden anhand von Überschrift, Abstract, Einleitung und Inhaltsübersicht schrittweise selektiert. Nur wenige Publikationen enthielten Informationen oder

S1	S2	S3	S4	S5	S6	S7	S8	S9
128/47	16/11	48/36	122/66	96/79	106/68	145/50	142/82	62/25

Tab. 1: Anzahl der Resultate nach Eingabe der Suchstrings (gesamt/ab 2012)

statistische Grundlagen zur Erstellung eines Medienkompetenzmodells oder Testszenarien für den Praxiseinsatz und konnten daher als relevant eingestuft werden. Die nach kompletter Sichtung der Arbeiten als wichtig identifizierte Literatur wurde einer Vorwärts- und Rückwärtssuche unterzogen. Literatur vor dem Jahr 2012 fand bei der vertiefenden Recherche Beachtung, wurde aber nicht in gleichem Maße systematisch durchsucht. Relevante Arbeiten wurden abschließend kategorisiert (s. Abb. 1).

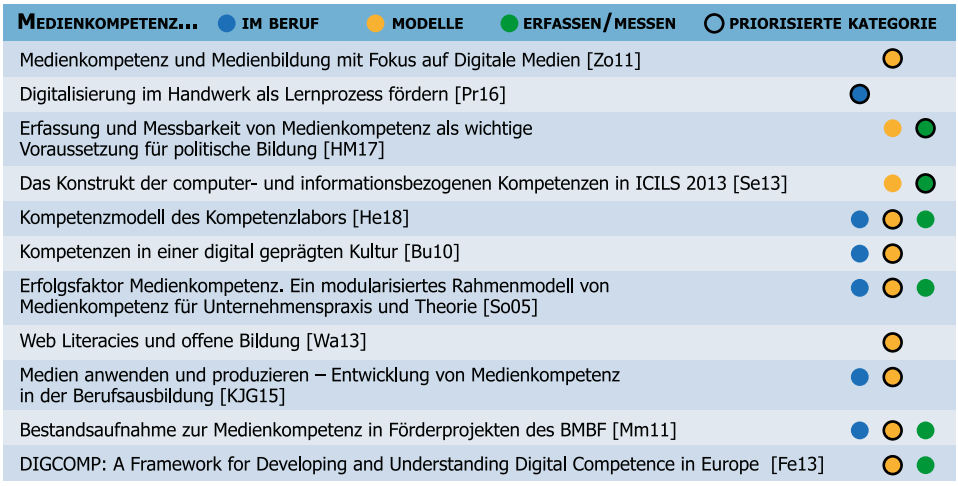


Abb. 1: Kategorisierung verwandter Arbeiten gemäß der empfohlenen Vorgehensweise von Webster und Watson [WW02]

2.1 Ergebnisse

Es zeigt sich, dass die Medienkompetenz nicht isoliert zu betrachten ist, sondern in Bezug zum Unternehmensumfeld und zu betrieblichen Arbeitsprozessen gesetzt werden muss [Pr16]. Dieser an der Praxis orientierte Blick auf die Medienkompetenz bringt einen starken Bezug zu digitalen Medien und Computern mit sich, der in vielen älteren Medienkompetenzmodellen – auch bedingt durch den schnellen technologischen Wandel – keine Beachtung findet [Zo11]. Durch diese Schnelllebigkeit ist es notwendig, Aspekte bisheriger Modelle auf ihre Aktualität zu prüfen, diese um neue Kompetenzen zu erweitern [So05] und gemäß der betrieblichen Arbeitsabläufe zu formulieren [He18]. Um die Medienkompetenz erfassen und messen zu können, sollte eine Struktur von Kompetenzbereichen, Kompetenzaspekten und Aufgaben mit unterschiedlichem Schwierigkeitsgrad zum Einsatz kommen [HM17].

Wichtige Kompetenzbereiche, -aspekte und Fertigkeiten liefert das Modell des EU-Projekts DIGCOMP [Fe13] sowie die Modelle nach Sohn [So05] und des Projekts „Medien anwenden und produzieren“ [KJL15]. Ebenso finden sich wichtige Elemente der Medienkompetenz in der Bestandsaufnahme zur Medienkompetenz in BMBF-Projekten [MM11], in den Empfehlungen der BMBF-Expertenkommission [BM10] sowie im Medienkompetenzmodell des Kompetenzlabors [He18] und in der Arbeit von Wittenbrink und Ausserhoffer [WA13]. Zur Erfassung der Medienkompetenz ist die Kombination verschiedener Methoden sinnvoll [MM11]. Möglich sind hierbei non-interaktive Tests (z. B. Multiple-Choice-Aufgaben), Performanceaufgaben (Software oder Computeranwendungen nutzen) und Autorenaufgaben (Informationsprodukte erstellen) [Se13]. Ebenso wie das eigentliche Modell müssen auch die gewählten Aufgaben aufgrund der schnellen Weiterentwicklung der IT- und Medienlandschaft immer wieder auf ihre Aktualität hin überprüft werden [MM11].

3 Experteninterviews

Als Ergänzung und Validierung der Literaturrecherche werden nachfolgend aggregierte Ergebnisse begleitender Interviews vorgestellt. Ziel ist es, die Praxisrelevanz zu stärken und gleiche Forschungsarbeiten auf diesem Feld auszuschließen. Befragt wurden vier Experten des Fachgebietes. Diese wurden anhand ihrer maßgeblichen Publikationen, die sich entweder im Bereich der Medienkompetenz im allgemeinen bewegen oder Medienkompetenz im Kontext der beruflichen Bildung und speziell im Handwerk betreffen, sorgfältig ausgewählt, um die Qualität der Ergebnisse sicherzustellen.

Experten: Dr. Harald Gapski – Leiter Grimme Forschung am Grimme-Institut, Dr. phil. Jörg Neumann – Berufspädagoge und Leiter der Abteilung „Medienstrategien“ an der TU Dresden, Dr. Lutz Goertz – Kommunikationswissenschaftler und Leiter Bildungsforschung beim mmb Institut sowie Dipl.-Psych. Jan Spilski – Projektmanager und wissenschaftlicher Koordinator des „Center for Cognitive Science“ an der TU Kaiserslautern.

Folgende Hypothesen sollen im Verlauf der Telefoninterviews mittels Leitfaden und festgelegtem Fragebogen verifiziert werden:

- H1: Medienkompetenz ist eine wichtige Kompetenz für Arbeitnehmer, die durch die zunehmende Digitalisierung eine ähnlich starke Rolle wie beispielsweise Fach- und Sozialkompetenz spielt.
- H2: In vielen (beruflichen) Ausbildungen wird Medienkompetenz nicht in ausreichendem Maße vermittelt, um auf die Aufgaben, welche die neuen digitalen Medien mitbringen, vorbereitet zu sein.
- H3: In vielen Handwerksbetrieben hat die Medienkompetenz noch keinen hohen Stellenwert, allerdings wird Medienkompetenz auch in dieser Domäne immer wichtiger, beispielsweise beim Einsatz von Software zur Planung von Baustellen.

- H4: Medienkompetenz lässt sich am besten anhand von handlungsorientierten Beispielen vermitteln. Das heißt, Aufgaben und Inhalte orientieren sich an Fragestellungen und Problemen aus Alltag und Beruf. Problemlösungen sollen dabei möglichst selbsttätig erarbeitet werden.
- H5: Wenn eine initiale Medienkompetenz (Grundkenntnisse zur Bedienung eines Computers/Smartphones) vorhanden ist, eignet sich die Vermittlung per E-Learning. Die Nutzung von E-Learning-Systemen stellt gleichzeitig ein Training für die Medienkompetenz dar.
- H6: Es gibt keine (öffentlich zugänglichen) standardisierten Messverfahren im Bereich der Medienkompetenz.
- H7: Es gibt keine Systeme, welche die Medienkompetenz des Anwenders automatisiert anhand der Nutzung von Software beurteilen.

Weiter sollen bekannte Medienkompetenzmodelle, softwaregestützte Trainings zur Erlangung und Methoden zur Ermittlung von Medienkompetenz erfasst werden.

3.1 Ergebnis

Bei der Auswertung der aufgezeichneten und verschriftlichten Interviews wurden die aufgestellten Hypothesen überwiegend verifiziert (vgl. Tab. 2).

H1	H2	H3	H4	H5	H6	H7
komplett	komplett	teilweise	komplett	teilweise	komplett	komplett

Tab. 2: Übersicht über den Verifizierungsgrad der Hypothesen

Eine Hypothese gilt als vollständig verifiziert, wenn alle Experten der Aussage uneingeschränkt zustimmen konnten. H3 wurde nur teilweise verifiziert, da alle Experten angaben, dass Medienkompetenz in Handwerksbetrieben bereits jetzt eine bedeutende Rolle innehat, diese aber in Zukunft noch zunehmen wird. Ebenso wurde auch H5 nur teilweise verifiziert. Alle Experten waren sich einig, dass sich bestimmte Bereiche gut per E-Learning vermitteln lassen, dies aber nicht für die Medienkompetenz in ihrer Gesamtheit gilt. Empfohlen werden deshalb eine gute tutorielle Begleitung und ein Blended-Learning-Konzept.

Weiter wurden die Medienkompetenzmodelle des EU-Projekts DIGCOMP [Fe13] und des Kompetenzlabors [He18] als relevante Arbeiten erfasst. Konkrete E-Learning-Angebote zur Erlangung einer umfassenden Medienkompetenz waren den befragten Experten nicht bekannt. Empfohlen wurde eine Ermittlung der Medienkompetenz durch eine Kombination verschiedener Testverfahren. Dabei sollten zum einen handlungsorientierte Aufgaben, aber auch Wissenstest sowie Selbst- und Fremdeinschätzungen eine wichtige Rolle spielen.

Damit wurde das Ziel der Interviews, die Informationen aus der Literaturrecherche mit aktuellen Informationen von Experten zu ergänzen, erfüllt. Die Betonung der Wichtigkeit der Medienkompetenz für Arbeitnehmer und der Umstand, dass sie bisher – trotz ihrer enormen Wichtigkeit – nicht ausreichend in die den Experten bekannten Ausbildungen Einzug findet, belegen außerdem die Relevanz dieser Arbeit.

4 Entwicklung des domänenspezifischen Medienkompetenzmodells

Als Grundlage des domänenspezifischen Medienkompetenzmodells wird zunächst ein generisches Modell entwickelt. Hierfür dient das im Verlauf der Recherche gefundene vollständigste Medienkompetenzmodell als Basis. Alle weiteren erfassten Medienkompetenzmodelle werden in ihre einzelnen Aspekte segmentiert, thematisch zusammengefasst und mit dem grundlagenbildenden Modell zusammengeführt. Durch Anpassen, Erweitern und Entfernen einzelner Kompetenzaspekte und Fertigkeiten wird das domänenspezifische Modell entwickelt, welches speziell für Meister des Stuckateurhandwerks gültig ist. Dieses spezifische Modell wird anschließend im Rahmen eines Workshops mittels Fragebogen evaluiert und, wenn nötig, angepasst.

4.1 Grundlage

Als Ausgangspunkt des Medienkompetenzmodells dient das Modell des EU-Projekts DIGCOMP [Fe13], da es das im Vergleich dieser Arbeit vollständigste Modell darstellt. Des Weiteren wurden Elemente des Modells nach Sohn [So05] und des Modells des Projekts „Medien anwenden und produzieren“ [KJL15] berücksichtigt und integriert, da sich diese durch einen starken Praxisbezug auszeichnen. Ebenso nahmen die Aspekte der Medienkompetenz, welche im Rahmen der Bestandsaufnahme der Medienkompetenz in BMBF-Projekten ermittelt wurden, [MM11] Einfluss. So wurden gleichzeitig die Empfehlungen der BMBF-Expertenkommission [BM10] zum Thema Medienkompetenz mit einbezogen. Außerdem finden sich Elemente der Modelle des Kompetenzlabors [He18] als eines der neuesten Modelle wieder. Auch die Arbeit von Wittenbrink und Ausserhoffer [WA13] beeinflusst die Entwicklung des Modells, da sich ihr Modell besonders mit dem Thema Web und digitale Medien beschäftigt. Auf eine explizite Einbeziehung der Modelle nach Baacke, Aufenanger, Groeben, Treumann und Pietras, die in der Literatur des Öfteren Erwähnung finden, wurde explizit verzichtet. Diese Modelle nehmen nur selten Bezug auf technikbezogene Kompetenzen und digitale Medien [Zo11], sind bereits bei der Erstellung verwendeter Modelle beachtet worden und nehmen dadurch Einfluss auf diese Arbeit.

4.2 Aufbau

Das entwickelte Medienkompetenzmodell soll eine Voraussetzung für die Messung von Medienkompetenz schaffen. Gemäß der Empfehlung von Martin und Herzig werden auf der

ersten Ebene Kompetenzbereiche (Dimensionen der Medienkompetenz) definiert, welche sich in verschiedene Aspekte aufgliedern. Anschließend werden konkrete Fertigkeiten mit unterschiedlichen Schwierigkeitsgraden beschrieben, um später konkrete Aufgaben daraus ableiten zu können. Dabei wurden die drei Fertigkeitsstufen des Modells von DigComp beibehalten [Fe13]. In Stufe Eins, den Anfängerfertigkeiten, werden ein generelles Problembewusstsein und Verständnis, sowie minimale Fertigkeiten vorausgesetzt. Die Fortgeschrittenenfertigkeiten beinhalten Sicherheit im Umgang mit den Technologien sowie ein tiefergreifendes Verständnis. Zuletzt steht die Expertenstufe für umfassendes Verständnis und umfassende Fertigkeiten. Durch die Aufschlüsselung der Kompetenzen werden diese beobachtbar, messbar und beurteilbar gemacht [HM17], [SS16].

4.3 Inhalte

Aufbauend auf den in Kapitel 4.1 erwähnten Modellen wurde ein neues Medienkompetenzmodell entwickelt, das eine generische Medienkompetenz beschreiben soll. Dieses Modell umfasst zunächst sieben Dimensionen:

- **Grundlagen**: grundlegendes Verständnis des Umgangs mit Medien
- **Informations- und Datenkompetenz**: Suche, Umgang und Verwaltung von Daten und Informationen
- **Kommunikation und Kollaboration**: interne und externe Kommunikation und Zusammenarbeit
- **Digitale Inhalte**: Erstellung von digitalen Inhalten und Umgang mit Software
- **Sicherheit**: Schutz von Endgeräten, personenbezogenen Daten und Privatsphäre
- **Problemlösung**: Lösung von Problemen mithilfe digitaler Technologien
- **Rahmenbedingungen**: rechtliche und ethische Aspekte

Das komplette Modell wurde anschließend anhand folgender Hypothesen auf seine domänenspezifische Zweckhaftigkeit geprüft und bei Bedarf angepasst, erweitert oder entfernt:

- Die Dimension/der Aspekt/die Fertigkeit ist wichtig, um den Beruf des Stuckateurs unter Einbeziehung neuer digitaler Medien ausüben zu können.
- Die Dimension/der Aspekt/die Fertigkeit ist wichtig, um einen Stuckateurbetrieb unter Einbeziehung neuer digitaler Medien führen zu können.

Evaluation. Das so für die Meisterausbildung im Stuckateurhandwerk adaptierte Medienkompetenzmodell wurde im Rahmen eines internen Workshops vorgestellt und evaluiert. Das

Teilnehmerfeld (n=6) der Befragung setzte sich aus allen Projektpartnern des Verbundprojekts D-MasterGuide zusammen. Ziele der Evaluation waren die Überprüfung der Relevanz einzelner Kompetenzen im Rahmen der Meisterausbildung im Stuckateurhandwerk und die Prüfung des Modells auf Vollständigkeit. Die Evaluation erfolgte in Form eines Fragebogens, bei dem die Teilnehmer die Relevanz der einzelnen Aspekte des Modells beurteilten und Anmerkungen machen konnten. Zudem wurden nicht berücksichtigte Kompetenzen abgefragt. Die vorgestellten Aspekte und Fertigkeiten wurden von den Teilnehmern der Befragung durchgehend als relevant eingestuft. Bei fünf Aspekten wurden kleinere Ergänzungen oder Anpassungen vorgeschlagen, die anschließend in das Modell übernommen wurden. Als fehlende Kompetenzen des Modells wurden mehrfach sogenannte weiche Kompetenzen aufgeführt, die als neue Dimension ergänzt wurden. Sie betreffen den persönlichen Umgang und die Einstellung zu digitalen Medien.

4.4 Medienkompetenzmodell für die Meisterausbildung im Stuckateurhandwerk

Resultierend aus der Evaluation ergibt sich abschließend ein adaptiertes Medienkompetenzmodell für das Stuckateurhandwerk in acht Dimensionen (s. Abb. 2).

Dimension 1: Grundlagen

1.1 IT-Verständnis
1.2 IT-Probleme lösen
1.3 Hardware
1.4 Wirtschaftliche/ökonomische Aspekte

Dimension 2: Informations- und Datenkompetenz

2.1 Suchen und Filtern von Daten, Informationen und digitalen Inhalten
2.2 Analyse und Reflexion von Daten, Informationen und digitalen Inhalten
2.3 Verwalten von Daten, Informationen und digitalen Inhalten

Dimension 3: Kommunikation und Kollaboration

3.1 Interaktion durch digitale Technologien
3.2 Teilen und Kollaboration durch digitale Technologien
3.3 Digitalisierung administrativer Aufgaben
3.4 Unternehmensprofil in digitalen Medien
3.5 Kommunikationsregeln

Dimension 4: Erstellung digitaler Inhalte

4.1 Digitale Inhalte
4.2 Einrichtung digitaler Tools zur Inhaltserstellung

Dimension 5: Sicherheit

5.1 Schutz von Endgeräten und Online-Accounts

Dimension 6: Problemlösung

6.1 Bedürfnisse und technologische Lösungen identifizieren
6.2 Prozesse im Betrieb verstehen und technisch beurteilen
6.3 Kleine bekannte Referenzlösungen einführen

Dimension 7: Rahmenbedingungen

7.1 Urheberrechte und Lizenzen
7.2 Datenschutz
7.3 Ethische Aspekte

Dimension 8: Weiche Kompetenzen

8.1 Offenheit für Veränderung und Experimentierfreude
8.2 Initiative und Lernbereitschaft

Abb. 2: Überblick über die Inhalte des Medienkompetenzmodells für das Stuckateurhandwerk

Wie konkrete Aspekte im Modell und die Steigerung der einzelnen Fertigkeitsstufen aussehen können wird in Abb. 3 veranschaulicht.

Das vollständige Modell ist online unter bit.ly/2QyP8J9 abrufbar. Es zeigt sich, dass es nur wenige konkrete berufsspezifische Kompetenzen gibt. Vielmehr zeichnet sich das Modell dadurch aus, dass durch die verschiedenen Fertigkeitsstufen auch sehr gering ausgeprägten

Dimension 2: Informations- und Datenkompetenz

1.1. Suchen und Filtern von digitalen Daten, Informationen und Inhalten

(1) Ich kann mithilfe einer Suchmaschine online nach Informationen suchen.
(2) Ich kann verschiedene Suchmaschinen nutzen, um nach Informationen zu suchen. Ich benutze Filter bei der Suche (z. B. nur Bilder, Videos oder Karten suchen).
(3) Ich kann fortgeschrittene Suchstrategien anwenden (z. B. Suchoperatoren) um die Suchanfrage im Internet einzugrenzen.

Abb. 3: Auszug aus dem Medienkompetenzmodell für das Stuckateurhandwerk

Kompetenzen, die heute oft automatisch impliziert werden, geprüft werden können. Dazu kommen Kompetenzaspekte, die für die Führung eines kleinen Unternehmens notwendig sind, informatische Kompetenzen wiederum entfallen teilweise.

5 Medienkompetenzerfassung innerhalb von E-Learning-Systemen

Das entwickelte Modell soll in der Praxis bei der Erfassung in E-Learning-Systemen eingesetzt werden. Hierfür werden mögliche Methoden zur Erfassung von Medienkompetenz und ein Bepunktungsschema für die erreichte Medienkompetenz vorgestellt.

5.1 Methoden zur Erfassung

Wie im Rahmen der Recherche empfohlen, sollte eine Mischung aus verschiedenen Methoden eingesetzt werden, um Medienkompetenz zu erfassen [MM11]. Die verschiedenen Methoden lassen sich grob in drei Kategorien einordnen: Tests und Aufgaben, Selbst-/Fremdeinschätzung und Erfassung mittels Systemnutzung. Die Aufgaben unterteilen sich in konvergente und divergente Aufgabentypen [Mc02]. Wie von den Experten empfohlen, sollten handlungsorientierte Aufgaben eingesetzt werden. Das stellt einen „[. . .] direkte[n] Bezug zu den konkreten Handlungssituationen des Lernenden in seinem Arbeitsfeld“ her [Br05]. Demnach können folgende Möglichkeiten genutzt werden:

Konvergente Aufgaben. Aufgaben, die sich durch eine genau definierte Lösung auszeichnen und deren Bewertung durch das System erfolgen kann [Gr10]. Sie eignen sich insbesondere für Aspekte, die stark mit Faktenwissen verknüpft sind, wie z. B. im Bereich Urheberrechte, Lizenzen und Datenschutz [GU11].

Divergente Aufgaben. Aufgaben, die auch Hintergrundwissen, Lösungswege und Begründungen im Rahmen von Freitextaufgaben, Tabellen, etc. erfassen können. Eigenständigkeit, Selbstvertrauen, Problembewusstsein und Flexibilität sollen gefördert werden. Die Aufgaben sollen „zu grundlegenden methodischen Überlegungen anregen, eine inhaltliche, qualitative Argumentation initiieren und damit die vertiefte Auseinandersetzung mit dem Lehrstoff bewirken“ [GU11]. Durch die Komplexität der Antwortmöglichkeiten ist eine manuelle Bewertung erforderlich. Aufgrund der vielfältigen Ausgestaltungsmöglichkeiten eignet

sich dieser Aufgabentyp zum Erfassen einer Vielzahl von Kompetenzen. So kann in einer Freitextaufgabe, die das Formulieren einer E-Mail erfordert, sowohl Fachwissen als auch Medienkompetenz im Bereich der „Kommunikationsregeln“ geprüft werden (z. B. ob die E-Mail einem korrekten inhaltlichen Aufbau folgt).

Selbsteinschätzung. Die Selbsteinschätzung ist insbesondere für weiche Kompetenzen, die nicht durch Tests geprüft werden können, wichtig. Die Überprüfung erfolgt in Form eines Fragebogens, für den handlungsorientierte Aussagen formuliert werden. Bei diesen kann angegeben werden ob oder inwieweit die Aussagen zutreffen.

Fremdeinschätzung. Die Fremdeinschätzung erfolgt analog zur Selbsteinschätzung, mit dem Unterschied, dass ein Lehrender (der den Lernenden beispielsweise in einem Flipped-Classroom-Konzept beobachten konnte) den Fragebogen ausfüllt.

Erfassung anhand der Systemnutzung. Durch die Fähigkeit des Lernenden das E-Learning-System aktiv zu nutzen, wird automatisch ein Nachweis für Teile der Medienkompetenz erbracht. Beispielsweise kann die Registrierung mit dem Aspekt „Schutz von Endgeräten und Online-Accounts“ verknüpft werden. Hierfür sollten im System entsprechende Mechanismen verankert werden, sodass beispielsweise zwingend ein sicheres Passwort erstellt werden muss. Gleichzeitig erhält der User an dieser Stelle Wissen über Passwortsicherheit. Nach der Registrierung können dann Kompetenzen im entsprechenden Bereich anerkannt werden. Dabei gibt es fließende Übergänge zwischen den einzelnen Methoden. So kann bei der Lösung divergenter Aufgaben auch eine Fremdeinschätzung der Medienkompetenz erfolgen. Ebenfalls kann bei der Bearbeitung von Aufgaben, z. B. bei der Bearbeitung einer Tabelle, eine Erfassung von Medienkompetenz durch das System stattfinden.

5.2 Ausblick Messung im System und Einsatz von Recommendern

Die im System erfassten Kompetenzen können nun einem persönlichen Kompetenzprofil zugewiesen werden. Hierfür werden einzelne Aspekte in handlungsorientierte Fertigkeiten unterteilt. Diese können bei Bedarf weiter in Subpunkte untergliedert werden. Die Subpunkte werden als vorhanden erfasst, sobald sie an einer Stelle im System nachgewiesen werden. Die Fertigkeit „Ich halte mich an Regeln bei der mündlichen und schriftlichen Kommunikation“ kann z. B. in die Subpunkte angemessene Begrüßung, Rechtschreibung, Aufbau von Briefen und Netiquette gegliedert werden. Sobald der Lernende nachweist, dass er Briefe formal korrekt aufbauen kann, wird ein Punkt für den entsprechenden Subpunkt angerechnet.

Wie viele Subpunkte jeweils benötigt werden, ist von der konkreten Fertigkeit abhängig. Wird bei komplexen Fertigkeiten eine Wiederholung als wichtig erachtet, können den entsprechenden Subpunkten mehrere Wiederholungen zugeordnet werden.

Da sich die Anzahl der Subpunkte innerhalb der Fertigkeiten, und somit auch die Menge der Aspekte und Dimensionen stark unterscheiden können, entsteht eine heterogene Verteilung, was bei der Betrachtung des Gesamtscores Beachtung finden muss [Ba11]. Um diesen Effekt

auszugleichen, erfolgt die Bewertung prozentual. Dadurch wird gleichzeitig die Möglichkeit geschaffen, Aspekte und Fertigkeiten unterschiedlich stark zu gewichten. Dies ist Teil der zukünftigen Projektarbeit. Eine grafische Darstellung des Bewertungsschemas ist in Abb. 4 zu sehen. Die Prozentzahl stellt den Anteil am übergeordneten Aspekt dar.

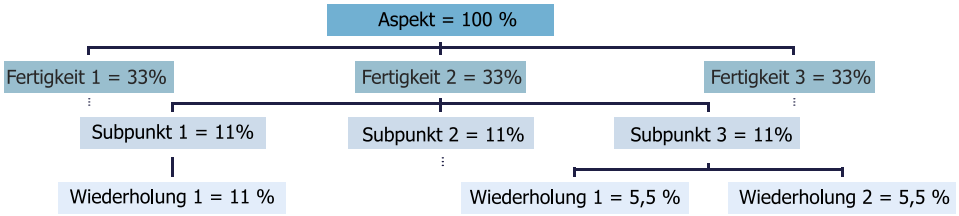


Abb. 4: Bewertungsschema für die Erfassung innerhalb des E-Learning-Systems

Durch die Erfassung im System wird der Einsatz eines Recommenders möglich. Dieser prüft, welche Kompetenzen noch erlangt werden müssen, welchen Lernpfad Nutzer mit ähnlicher Kompetenz genommen haben, und welcher Lernpfad der effizienteste war [KT04]. Darauf aufbauend werden Empfehlungen für verfügbare Lernressourcen abgegeben, um den Lernenden bei der Erlangung von Medienkompetenz zu unterstützen.

6 Fazit

Sowohl die Literaturrecherche als auch die Expertenbefragungen haben die Annahme, dass die Medienkompetenz durch die zunehmende Digitalisierung eine immer wichtigere Rolle für Arbeitnehmer einnimmt, bestätigt. Es zeigte sich, dass in der beruflichen Ausbildung innerhalb der untersuchten Domäne zumeist keine ausreichende Vermittlung dieser stattfindet. Bisher fehlt ein Messverfahren, um die Medienkompetenz in ihrer Gesamtheit innerhalb eines E-Learning-Systems zu erfassen. Ebenso fehlt ein umfassendes Kompetenzmodell, das die Grundlage der Messung bilden kann.

Ziel der Arbeit war die theoretische Konzeption eines Rahmenwerkes, um Medienkompetenz in E-Learning-Systemen erfassen zu können und somit eine Medienkompetenzvermittlung möglich zu machen. Dieses Ziel wurde durch die Entwicklung eines umfassenden Medienkompetenzmodells erreicht. Durch den Einsatz der Informationen aus der Literaturrecherche und den Expertenbefragungen wurde ein Modell geschaffen, welches die Medienkompetenz in ihrer Gesamtheit beschreibt. Die konkreten handlungsorientierten Fertigkeiten, die den Aspekten der Medienkompetenz zugeordnet sind, stellen einen hohen Praxisbezug des Modells sicher. Aufbauend auf dem entstandenen Modell wurden erste Ansätze entwickelt, um die entsprechenden Kompetenzen innerhalb eines E-Learning-Systems zu messen. Mögliche Methoden aus den Bereichen Tests und Aufgaben, Selbst- und Fremdeinschätzung, sowie Erfassung mittels Systemnutzung wurden dargestellt. Vor allem die Erfassung der Medienkompetenz während und anhand der Systemnutzung birgt noch viele Potenziale für die Zukunft. So wäre es beispielsweise denkbar, Logfiles auszuwerten, um die Medienkompetenz der Anwender zu beurteilen.

Kritisch zu sehen ist, dass die Vollständigkeit eines Modells nie als gesichert betrachtet werden kann. Insbesondere digitale Technologien unterliegen einem schnellen Wandel, der immer neue Kompetenzen fordert. Deshalb ist es zwingend notwendig, das Modell immer wieder auf seine Aktualität hin zu überprüfen. Außerdem ist eine genaue Messung von Medienkompetenz trotz eines umfassenden Kompetenzmodells nur schwer durchführbar. Speziell weiche Faktoren, die den persönlichen Umgang und die Einstellung gegenüber digitalen Medien beschreiben, sind schwer und nur indirekt messbar.

Das entwickelte Modell ermöglicht eine Adaption für weitere Bereiche, sofern eine genaue Überprüfung der benötigten Kompetenzen vorgenommen wird, sodass auch Forschungsarbeiten aus anderen Domänen darauf aufbauen können. Modelle, Fragebogen und Rahmenwerk stehen für diesen Zweck öffentlich zur Verfügung: http://bit.ly/2Fas3Z4. Das genutzte Vorgehen und die resultierenden Modelle bieten somit eine solide Basis für weitere Forschung in der E-Learning basierten Medienkompetenzvermittlung.

Literatur

[Ba11] Balceris, M.: Medien- und Informationskompetenz - Modellierung und Messung von Informationskompetenz bei Schülern./, 2011.

[Bi16] Bitkom e.V.: Neue Arbeit - Digitalisierung schafft neue Jobs für Fachkräfte./, 2016.

[BM10] BMBF: Kompetenzen in einer digital geprägten Kultur./, 2010.

[BM16] BMBF: Bildungsoffensive für die digitale Wissensgesellschaft./, 2016.

[BM17] BMBF: eQualification 2017 – Lernen und Beruf digital verbinden./, 2017.

[BM19] BMBF, 2019, URL: https://www.qualifizierungdigital.de/de/projektdatenbank-27.php?D=168&F=0&FS=czo2OiJwYWRpZ2kiOw%3D%3D&M=445&T=1.,25.04.2019., Stand: 19. 06. 2019.

[Br05] Bremer, C.: Handlungsorientiertes Lernen mit Neuen Medien. In: Online-Pädagogik, Band 2. B. lehmann und E. Bloh, S. 175–197, 2005.

[Br09] v. Brocke, J.; Simons, A.; Niehaves, B.; Reimer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process. In: ECIS. 2009.

[EB15] Eichhorst, W.; Buhlmann, F.: Die Zukunft der Arbeit und der Wandel der Arbeitswelt./, 2015.

[EB19] EBusiness-KompetenzZentrum GUG, 2019, URL: http://d-masterguide.de/, Stand: 19. 06. 2019.

[Eu06] Europäisches Parlament und Rat der Europäischen Union: Empfehlung des Europäischen Parlaments und des Rates vom 18. Dezember 2006 zu Schlüsselkompetenzen für lebensbegleitendes Lernen. Amtsblatt der Europäischen Union L/394, S. 10–18, 2006.

[Fe13] Ferrari, A.: DIGCOMP: A Framework for Developing and Understanding Digital Competence in Europe./, 2013.

[Gr10] Gruttmann, S.: Formatives E-Assessment in der Hochschullehre: computerunterstützte Lernfortschrittskontrollen im Informatikstudium./, 2010.

[GU11] Gruttmann, S.; Usener, C. A.: Prüfen mit Computer und Internet. Didaktik, Methodik und Organisation von E-Assessment. In: L3T - Lehrbuch für Lernen und Lehren mit Technologien. S.Schön und M. Ebner, 2011.

[He18] Helliwood media & Education: Kompetenzmodell des Kompetenzlabors./, 2018.

[HM17] Herzig, B.; Martin, A.: Erfassung und Messbarkeit von Medienkompetenz als wichtige Voraussetzung für politische Bildung. In: Medienkompetenz - Herausforderung für Politik, politische Bildung und Medienbildung. H. Gapski und M. Oberle, S. 125–135, 2017.

[KJL15] Krämer, H.; Jordanski, G.; L.Goertz: Medien anwenden und produzieren – Entwicklung von Medienkompetenz in der Berufsausbildung./, 2015.

[KT04] Koper, R.; Tattersall, C.: New directions for lifelong learning using network technologies. British Journal of Educational Technology 36/, S. 689–700, 2004.

[Mc02] McAlpine, M.: Principles of Assessment./, 2002.

[MM11] MMB-Institut für Medien- und Kompetenzforschung: Bestandsaufnahme zur Medienkompetenz in Förderprojekten des BMBF./, 2011.

[Pr16] Prescher, T.; Hellriegel, J.; Schön, M.; Baumann, A.; Heil, M.; Schulz, F.: Digitalisierung im Handwerk als Lernprozess fördern. In: Proceedings of DeLFI Workshops 2016. R. Zender, S. 209–215, 2016.

[Sc09] Schorb, B.: Gebildet und kompetent. Medienbildung statt Medienkompetenz. merz. 2009/05, S. 50–56, 2009.

[Se13] Senkbeil, M.; Goldhammer, F.; Bos, W.; Eichelmann, B.; Schwippert, K.; Gerick, J.: Das Konstrukt der computer- und informationsbezogenen Kompetenzen in ICILS 2013. In: ICILS 2013. W. Bos u. a., S. 82–112, 2013.

[So05] Sohn, M.: Erfolgsfaktor Medienkompetenz. Ein modularisiertes Rahmenmodell von Medienkompetenz für Unternehmenspraxis und Theorie./, 2005.

[SS16] Sauter, W.; Staudt, F.-P.: Strategisches Kompetenzmanagement 2.0. Springer Fachmedien, 2016.

[WA13] Wittenbring, H.; Ausserhofer, J.: Web Literacies und Offene Bildung. In: Netzpolitik in Österreich. C. Landler, P. Parycek und M. C. Kettemann, S. 225–235, 2013.

[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.

[Zo11] Zorn, I.: Medienkompetenz und Medienbildung mit Fokus auf Digitale Medien. MedienPädgogik/20, S. 175–209, 2011.

Chancen und Herausforderungen von Virtual Reality in der Aus- und Weiterbildung im Gesundheitswesen

Julian Schuir[1], Alina Behne[1] und Frank Teuteberg[1]

Abstract: Virtual Reality hat in den vergangenen Jahren zunehmend an Aufmerksamkeit gewonnen. Vor allem innerhalb der Aus- und Weiterbildung im Gesundheitswesen werden in der Forschung zunehmend Lösungen entwickelt und implementiert. In der Praxis stellt der Einsatz virtueller Lernumgebungen derzeit noch eine Ausnahme dar. Vor diesem Hintergrund verfolgt der vorliegende Beitrag das Ziel, die Chancen und Herausforderungen des Einsatzes von Virtual Reality zur Aus- und Weiterbildung im Gesundheitswesen mittels einer STEP-Analyse näher zu beleuchten. Die Ergebnisse einer Literaturrecherche und zuvor durchgeführter Interviews mit Gesundheitsexperten zeigen, dass insbesondere von einer erhöhten Patientensicherheit sowie von besseren Lernergebnissen profitiert werden kann. Demgegenüber stehen Herausforderungen, zu denen fehlende, schulungsbasierte Softwareanwendungen sowie mangelnde IT-Kompetenzen aufseiten der Lehrenden gehören.

Keywords: Virtual Reality, Aus- und Weiterbildung, Gesundheitswesen, STEP-Analyse

1 Einleitung

Seit der Ankündigung des Oculus Rift Developer Kits im Jahr 2013 gewann Virtual Reality (VR) zunehmend an Aufmerksamkeit. Sinkende Preise und erhebliche technologische Verbesserungen sorgten für steigende globale Absatzzahlen von Virtual-Reality-Headsets und einem wachsenden Interesse an dieser Technologie [JK18]. Auch für die Zukunft wird VR eine steigende wirtschaftliche Bedeutung zugewiesen: Einer Prognose von Goldman Sachs zufolge werden globale VR-Erlöse bis zum Jahr 2025 den TV-Markt überholen [Go16]. Besonders der Einsatz von VR in der Aus- und Weiterbildung wird vielfach diskutiert [FO15]. Ein wiederkehrendes Argument hierfür stellt bspw. die Möglichkeit zur wiederholenden Erprobung von prozeduralen Fähigkeiten in einer realitätsnahen, nicht-gefährlichen Umgebung dar [Fo18]. Vor diesem Hintergrund prognostizierte der Technologie-Report des New Media Consortium aus dem Jahr 2016 eine Diffusion der virtuellen Realität zu Bildungszwecken im Laufe der Jahre 2018 und 2019 [Ne16].

Bei der Beantwortung der Frage, wie Virtual Reality die Bildung verändern kann, beschränken sich viele Autoren primär auf die vereinfachte Darstellung komplexer Zusammenhänge und damit auf den Lernerfolg [Co12]. Daraus resultiert der Bedarf, die Vor- und Nachteile des Einsatzes von VR als Bildungsmedium tiefergehend zu untersu-

[1] Universität Osnabrück, Fachgebiet Unternehmensrechnung und Wirtschaftsinformatik, Katharinenstr. 1, 49076 Osnabrück, {julian.schuir; alina.behne; frank.teuteberg}@uni-osnabrueck.de

doi:10.18420/inf2019_81

chen. Hierzu sollen die Chancen und Herausforderungen aus Sicht der Forschung sowie weiterer involvierter Stakeholder ermittelt werden.

Ein in der Forschung verbreitetes und in Deutschland sehr relevantes Einsatzgebiet stellt die Aus- und Weiterbildung klinischer Mitarbeiter wie Medizinern, Pflegenden und weiteren Akteuren des Gesundheitswesens dar. Im Jahr 2018 waren über 5,5 Millionen Menschen in deutschen Krankenhäusern beschäftigt [St17]. Dabei hat sich die medizinische Aus- und Weiterbildung in den vergangenen Jahren kontinuierlich weiterentwickelt. Das klassische Lehrkonzept „See one, do one, teach one“ [Gu15], das mit der Erprobung von praktischen Kompetenzen am Patienten einhergeht, ist aufgrund ethischer und wirtschaftlicher Bedenken nicht länger praktikabel. Massive strukturelle Veränderungen wie der Wandel von der stationären hin zur ambulanten Medizin erfordern effiziente Behandlungsmethoden. Mediziner, Pflegende und weitere klinische Mitarbeiter müssen sich kontinuierlich weiterbilden, um neuen Forschungserkenntnissen und den daraus resultierenden innovativen Behandlungsmethoden und -instrumenten gerecht zu werden [Gu15]. Aus diesen Überlegungen wird die folgende Forschungsfrage abgeleitet, die in diesem Beitrag adressiert wird:

Welche Chancen und Herausforderungen sind mit der Nutzung von Virtual Reality als Bildungsmedium im Gesundheitswesen verbunden?

Der Beitrag gliedert sich wie folgt: Kapitel 2 skizziert die Charakteristika moderner VR-Systeme und ihre Eignung zu Bildungszwecken im Gesundheitswesen. Das dritte Kapitel umfasst das methodische Vorgehen der vorliegenden Untersuchung bestehend aus einer Literaturrecherche und Experteninterviews. Daraufhin werden in Kapitel 4 die verwandten Arbeiten näher beleuchtet. Kapitel 5 legt anschließend die mit der Nutzung assoziierten Chancen und Herausforderungen mithilfe der STEP-Analyse dar. Abschließend werden die aus den Ergebnissen resultierenden Implikationen für die Forschung sowie die Wirtschaft diskutiert.

2 Virtual Reality als Bildungsmedium

VR bezeichnet die Nutzung von dreidimensionalen Computergrafiken, um das Gefühl der Immersion in eine interaktive, virtuelle Umgebung zu vermitteln [Pa06]. Hierzu werden sog. Head-Mounted-Displays (HMD) verwendet, die das Eintauchen in eine virtuelle Welt ermöglichen [OF15]. Die resultierende visuelle Erfahrung kann durch haptische und akustische Stimuli angereichert werden, um die Realitätsnähe zu erhöhen. Zur Interaktion mit der virtuellen Umgebung können Gesten, Sprache oder Controller dienen [Is18]. Abzugrenzen ist diese Form der VR von der sog. nicht-immersiven VR, die nicht über HMD, sondern über konventionelle Bildschirme abgebildet wird und vor allem in Computerspielen zu finden ist [OF15].

In ihrem systematischen Literaturreview identifizieren Mikropoulos und Natsis (2011) eine Reihe von lernfördernden Merkmalen der VR. Dazu zählen bspw. die Darstellung in

der Ego-Perspektive sowie das Gefühl der Präsenz, die beide mit der intensiven Erfahrung des Seins in der virtuellen Umgebung zusammenhängen [MN11]. Eine Motivation zur Nutzung von VR als Bildungsmedium liegt in der Darstellung von Szenarien, deren Durchführung in der Realität zu gefährlich, ethisch problematisch oder schwer zugänglich ist [Pa96].

Im Gesundheitswesen wird VR seit über 20 Jahren für Simulationen genutzt [Is18]. Simulationen bezeichnen künstliche Darstellungen klinischer Szenarien [Wo18]. Sie dienen der Imitation realer Patienten, der Darstellung anatomischer Regionen oder klinischer Aufgaben sowie der Veranschaulichung realistischer Umstände, in denen medizinische Dienstleistungen erforderlich sind [Gu15]. Bisher werden in der medizinischen Bildung primär analoge Simulationen eingesetzt, die sich auf die Darstellung einzelner, auf das Fallszenario zugeschnittener Sachverhalte konzentrieren und haptische Einzelanfertigungen wie z. B. Puppen beinhalten. Mit der Verwendung dieser Simulatoren gehen hohe Kosten und eine beschränkte Wiederverwendbarkeit für weitere Fallszenarien einher [Gu15], [Sa18].

VR hingegen bietet die Möglichkeit, reale Umgebungen wie z. B. Behandlungszimmer und medizinische Geräte komplett virtuell abzubilden. Der immersive Charakter trägt während der Nutzung dazu bei, sich in der virtuellen Umgebung tatsächlich anwesend zu fühlen [Gu15]. Ein Beispiel hierfür stellt die von Wong et al. (2018) in ihrem Beitrag vorgestellte Applikation dar. Sie dient zur Übung von Reanimationsmaßnahmen mithilfe eines Defibrillators. Die Ausgangssituation bildet in diesem Fall eine Person, die regungslos am Boden liegt. Zur Kompetenzvermittlung werden dem Nutzer im ersten Modus die einzelnen Schritte einer Reanimationsmaßnahme am Patienten vorgeführt. Ein Avatar führt diese im virtuellen Raum vor. Im zweiten Modus muss der Benutzer die einzelnen Schritte unter Berücksichtigung einer zeitlichen Beschränkung selbst ausführen, indem er einen virtuellen Defibrillator in der künstlichen Umgebung bedient [Wo18].

3 Methodik

Um die Chancen und Herausforderungen von VR als Bildungsmedium zu identifizieren, wurde in diesem Beitrag ein multimethodisches Vorgehen verwendet. Im ersten Schritt wurde eine Literaturrecherche beruhend auf dem Vorgehen nach vom Brocke et al. (2009) durchgeführt [Vo09]. Im Fokus der Literatursuche standen Artikel, die frühestens im Jahr 2013 erschienen sind und die aktuelle Generation der VR-Hardware als Bildungsmedium im Gesundheitswesen adressieren. Hierzu wurden die Datenbanken EBSCOhost, ScienceDirect, SpringerLink, IEEE und Pubmed mithilfe des folgenden Suchstrings durchsucht: *(„virtual reality "OR „head mounted display") AND (health OR healthcare OR medicine) AND education.* Nachfolgend wurden die Beiträge anhand der Titel, Abstracts und Volltexte geprüft. Insgesamt wurden 26 Beiträge als relevant erachtet und bilden die Basis der Literaturanalyse. Zusätzlich wurde auf Basis der identi-

fizierten Literatur eine Vorwärts- und Rückwärtssuche durchgeführt, in der sieben weitere Beiträge identifiziert wurden.

Im zweiten Schritt wurden im Zeitraum von Dezember 2018 bis Februar 2019 acht leitfadengestützte Experteninterviews mit Akteuren aus dem Gesundheitswesen durchgeführt und mithilfe einer qualitativen Inhaltsanalyse nach Gläser und Laudel (2010) ausgewertet [GL10]. Die Schlüsselfragen thematisieren Chancen und Herausforderungen sowie zukünftige Bedarfe von VR im Gesundheitswesen. Die Ergebnisse werden in Kapitel 4 gemeinsam mit den Resultaten der Literaturrecherche vorgestellt und fließen in die Diskussion in Abschnitt 5 ein. Tab. 1 zeigt eine Übersicht zu Beruf und Erfahrung der interviewten Experten sowie die Dauer der Interviews. Die Erfahrung wird ab dem Zeitpunkt summiert, zu dem die Befragten ihre berufliche Ausbildung im Gesundheitswesen begonnen haben.

Nr.	Beruflicher Hintergrund	Erfahrung (in Jahren)	Dauer (in Min.)
E1	Notdienstsanitäter	13	27
E2	Rettungssanitäter	10	21
E3	Medizinstudent	6	30
E4	Medizinstudent	6	25
E5	Medizinstudent	5	31
E6	Facharzt für Anästhesie und Notfallmedizin	19	38
E7	Facharzt für Orthopädie und Unfallchirurgie	25	41
E8	Oberarzt einer kinderonkologischen Station	30	28

Tab. 1: Übersicht der Experteninterviews

4 Verwandte Arbeiten

Innerhalb der vergangenen Jahre hat die Forschung zum Einsatz von VR zu Bildungszwecken in den einzelnen wissenschaftlichen Disziplinen stark zugenommen [JK18]. Neben der Informatik untersuchen vor allem die Ingenieurs-, und Sozialwissenschaften immersive Bildungsszenarien aus verschiedenen Perspektiven. Dabei wird insbesondere die Schulung von medizinischen Berufsgruppen wie Ärzten adressiert [OF15], [Sa18]. Die Untersuchungsschwerpunkte liegen in der Gestaltung von Artefakten [Bu18], [DSG18], [Lu13], [Sn19], [So17] sowie in der Wirkungsforschung [Gu18], [Ra16]. Auf die positiven und negativen Folgen des Einsatzes von aktueller VR-Hardware im Gesundheitswesen konzentrieren sich dabei nur wenige Autoren.

McGrath et al. (2017) untersuchen die Eignung von virtuellen Lernumgebungen in Form von Augmented Reality, Virtual Reality und Serious Games zur Kompetenzvermittlung im Gesundheitswesen. VR eigne sich demnach vor allem zur Simulation von Katastrophen und Großereignissen sowie zur Aneignung von vorgangsbezogenem Wissen. Auch regelmäßige Leistungskontrollen für diese Szenarien können in der virtuellen Umgebung abgebildet werden. Verbesserungsmöglichkeiten sehen die Autoren im Bereich des Realitätsgrades. Durch fehlende haptische Devices werde die Akzeptanz von VR erschwert. Ferner stellen die logistischen Anforderungen eines VR-Systems eine weitere Adoptionsbarriere dar [Mc18]. Iserson (2018) stellt in seinem Beitrag fest, dass die Nutzung von VR in der medizinischen Ausbildung Mehrwerte in Bezug auf die Patientensicherheit und das gesellschaftliche Ansehen von Ärzten hervorbringt, indem medizinische Fehler reduziert werden können. Hierzu schlägt der Autor vor, VR sowohl als Bildungs- und als Prüfungsmedium im medizinischen Bereich im frühen Ausbildungsstadium zu nutzen, bevor am Patienten selbst gearbeitet wird. Demgegenüber stehen kritische Gesichtspunkte wie hohe Investitionskosten [Is18]. Ähnlich dazu beleuchten Sambadeik et al. (2018) in ihrem Literaturreview die Vor- und Nachteile sowie Vorschläge zur Verbesserung der VR-Nutzung in der Bildung medizinischer Berufsgruppen. Die Autoren beschränkten sich auf Studien, die zwischen den Jahren 2012 und 2016 veröffentlicht wurden [Sa18].

Hervorzuheben ist, dass sich die genannten Autoren meist auf einzelne Aspekte wie z. B. ethische Vorteile oder die Patientensicherheit beziehen. Demgegenüber stehen Sambedeik et al. (2018), welche mehrere Vor- und Nachteile von VR in der medizinischen Bildung diskutieren, methodisch jedoch nur auf eine Literaturrecherche aufbauen. Generell ist bei den verwandten Arbeiten zu beachten, dass sich diese nicht ausschließlich auf den Einsatz von Virtual Reality im Sinne von HMDs zu Schulungszwecken konzentrieren, sondern der Begriff weiter gefasst wird. Der vorliegende Beitrag unterscheidet sich insofern von den zuvor erläuterten Beiträgen, da sich allein auf die Nutzung von HMD-basierten VR-Systemen konzentriert wird. Außerdem erfordert die Innovationsgeschwindigkeit im Segment der virtuellen Realität eine kontinuierliche Aufarbeitung der Forschungsergebnisse. Durch den Einbezug der Zielgruppe von VR als Bildungsmedium im Gesundheitswesen mithilfe von Experteninterviews werden im vorliegenden Beitrag zusätzlich die Meinungen potentieller Anwender berücksichtigt.

5 Chancen und Herausforderungen

Ziel ist es, die mit dem Einsatz von VR in der Aus- und Weiterbildung assoziierten Chancen und Herausforderungen näher zu beleuchten. Dabei wird Bezug auf die Informationen aus Literatur sowie Experteninterviews genommen. Die Befragten zeigten insgesamt großes Interesse und Hoffnung in Bezug auf die VR-Nutzung in der Ausbildung des Gesundheitswesens. Bis auf einen Experten konnten die Interviewten keine Erfahrungen mit VR innerhalb ihres Berufes sammeln, weshalb ihre Meinungen größtenteils auf Informationen aus dem beruflichen Umfeld, dem Internet oder Messen basieren.

Generell empfanden diese jedoch den Fortschritt im Gesundheitswesen in Bezug auf die Digitalisierung und insbesondere auf VR sehr langsam [E1], [E2], [E4], [E5], [E6].

Nachfolgend wird die STEP-Analyse[2] als Strukturierungsinstrument angewandt. Sie dient der Kategorisierung soziokultureller, technologischer, ökonomischer und politischer Faktoren, die den Einsatz von VR als Bildungsmedium beeinflussen [PN07]. Abb. 1 veranschaulicht die Ergebnisse der untersuchten Publikationen und der Interviews, eingeteilt in Chancen (C) und Herausforderungen (H) sortiert nach den Kategorien der STEP-Analyse.

		Aspekte	Quelle		C	H
			Literatur	Interviews		
Dimension	soziokulturell	Verbesserung der Anschaulichkeit von Lerninhalten durch Interaktivität	[All8], [Fo18], [Gu15], [Si18], [Wo18]	[E1], [E3], [E4], [E7]	●	
		Steigerung der Lerneffektivität	[Ma19] , [Mc18], [So17], [SL17]	[E2], [E3], [E5], [E7], [E8]	●	
		Erhöhung der Lerneffizienz	[Is18]		●	
		Steigerung des Erlebnis- und Aktivierungsgehaltes	[All8], [Bu15], [Ja16], [Lo16], [Mo17], [SL17], [Wo18]	[E5]	●	
		Erhöhung der Patientensicherheit	[Is18], [Mc18], [Sn19], [ZC19]	[E3], [E8]	●	
		Gesundheitliches Risiko	[Da14], [DSG18], [HP16], [Hu17], [JK18], [Mo17], [Sn19]	[E7]		●
		Fehlende IT-Kompetenzen von Lehrenden	[DSG18]	[E5]		●
	technologisch	Ausweitung der Flexiblität	[All8], [Bu18], [Kl14], [SF17], [Gu15],[Ma19], [Mc18], [Mo17]	[E1], [E4], [E7], [E8]	●	
		Ortsunabhängkeit des Lernens	[Is18], [Lo16], [Ma19], [Sn19]		●	
		Mobilitätsprobleme der Hardware	[Da14], [DSG18], [JK18]			●
		Eingeschränkte Barrierefreiheit	[Fo18], [Wo18]			●
		Niedriger Reifegrad der Technik	[Kl14], [Mc18]	[E2]		●
	wirtschaftlich	Sinkende Hardwarekosten	[DSG18], [JK18]		●	
		Reduktion des Einsatzes von Leichnamen und tierischen Kadavern	[Ki13], [Mo17]	[E3], [E4]	●	
		Verbesserung der Kosteneffizienz	[Kl14], [Lo16]	[E4]	●	
		Betreuungs- und Wartungsaufwand	[DSG18], [Mc18]	[E1]		●
		Fehlende Softwareangebote	[Gu16], [JK18], [Mc18]	[E8]		●
		Hohe Implementierungs- und Evaluationsaufwendungen	[DSG18], [Sa18]	[E6]		●
	pol.	Fehlende staatlich geförderte Projekte		[E6]		●

Abb. 1 STEP-Analyse der Chancen und Herausforderungen

5.1 Soziokulturelle Dimension

Aus soziokultureller Sicht kann von einer **verbesserten Anschaulichkeit von Lerninhalten durch die interaktive Darstellung in VR** profitiert werden. Nach dem Kon-

[2] Die STEP-Analyse wird auch PEST-Analyse genannt und besteht aus den folgenden Dimensionen: Social, Technological, Economical und Political (engl.).

struktivismus, einer lernpsychologischen Theorie, müssen Lernende die Inhalte individuell erfahren und fühlen, um sich die Fähigkeiten im Sinne ihrer eigenen Denkstrukturen anzueignen [Fo18]. Dadurch kann eine **höhere Lerneffektivität** erreicht werden [Sa18], [So17], [E2], [E7], [E8]. Studierende können in anatomischen Fächern die komplexen Beziehungen zwischen Organen besser nachvollziehen und „mithilfe einer Reise durch den eigenen Körper Bezüge schneller herstellen als beim bloßen Lesen eines Buches“ [E6]. Eigene Fehler können besser reflektiert werden, indem das Lernerlebnis aufgezeichnet und Statistiken zur individuellen Performancemessung herangezogen werden [E5].

Ein weiteres Potential liegt in der **Erhöhung der Lerneffizienz** durch eine Beschleunigung des Lernprozesses [Is18]. Insbesondere Prozesswissen (z. B. über Operationen, Wiederbelebungsmaßnahmen, kardiologische Eingriffe) erfordert eine intensive Aneignung der Abläufe, die mehrere Jahre dauern kann [Gu15]. Vor allem Arztassistenten, Zahnmediziner sowie Psychotherapeuten können durch eine schnellere Aneignung dieser Fähigkeiten aufgrund von wiederholenden Übungen in VR profitieren [Is18]. Der Einsatz von VR trägt außerdem zur **Steigerung des Erlebnis- und Aktivierungsgehaltes** bei. Eine erhöhte intrinsische Motivation von Lernenden stellen Moro et al. (2017) fest, die in ihrem Experiment den Einsatz von Tablets, Augmented Reality und VR zur Vermittlung von anatomischen Kenntnissen vergleichen [Mo17], [E5]. Begründet liegt dies vor allem in dem realitätsnahen Gefühl der Anwesenheit in der virtuellen Umgebung [Lo16], [Ja16], [Bu15]. Unterstützt wird dieser Aspekt durch die vollständige Absorbierung störender Einflüsse aus der realen Umgebung [SL17].

Vor dem Hintergrund der gestiegenen Bedeutung der Patientensicherheit stellen medizinische Fehler ein großes Risiko dar. Jeder dritte Tod in den vereinigten Staaten ist auf derartige Fehler zurückzuführen [MD16]. Diese können reduziert werden, indem praktische Kompetenzen, die in der Vergangenheit zunächst wiederholt am Patienten erprobt wurden, in VR angeeignet werden. Damit bietet die Nutzung von VR Potentiale zur **Erhöhung der Patientensicherheit** [Mc18]. Mit dieser Chance wird eine Erhöhung des gesellschaftlichen Ansehens der Akteure im Gesundheitswesen assoziiert [Is18]. Außerdem bietet der Einsatz von VR eine bessere Vorbereitung für die Praxis, weil mit der realitätsnahen 3D-Simulation das Gefühl vermittelt wird, schon einmal in der Situation gewesen zu sein. Auf diesem Weg kann bspw. der Einstieg als Assistenzarzt vereinfacht werden, da realer Stress bereits in der Ausbildung simuliert wird [E3], [E8].

Den soziokulturellen Chancen stehen auch Herausforderungen wie das **gesundheitliche Risiko** aufseiten der Nutzer gegenüber. Bei der Verwendung von stereoskopischen 3D-Displays kann die sog. Motion Sickness (dt. Simulationskrankheit) auftreten [DSG18]. Einhergehende Nebenwirkungen umfassen Gefühle von Übelkeit, Schwindel, Orientierungslosigkeit, Kopfschmerzen, Müdigkeit und Sehstörungen [DSG18], [Mo17], [E7]. Moro et al. (2017) stellen im Rahmen ihres Experimentes fest, dass ein Drittel der Teilnehmenden ähnliche Symptome wie eine verschwommene Sicht und Konzentrationsprobleme äußerten [Mo17]. Weitere assoziierte gesundheitliche Risiken umfassen Realitätsverluste und psychopathologische Effekte [Hu17]. Eine weitere Herausforderung

aufseiten der Lehrenden stellen aktuell **mangelnde IT-Kompetenzen für die Betreuung** der VR-Hardware dar [DSG18], [E5].

5.2 Technologische Dimension

Aus technologischer Sicht bietet VR eine **Ausweitung der Flexibilität** in Bezug auf die darzustellenden Szenarien im Vergleich zu den aktuell verwendeten Simulatoren, die auf einzelne Anwendungsszenarien zugeschnitten sind [Al18]. Situationen mit verschiedensten Komplikationen können gefahrenfrei erprobt werden [Ma19], [E4], [E7]. Hierzu zählen der Umgang mit risikobehafteten Patienten [Kl14] z. B. bei Verweigerung eines Krankentransports oder die Überlieferung von negativen Nachrichten wie die Diagnose einer unheilbaren Krankheit [E1], [E8].

Des Weiteren kann durch VR eine **ortsunabhängige Verfügbarkeit** von Lerninhalten ermöglicht werden. Studierende der Medizin können ihre Fähigkeiten mithilfe von portablen Stand-Alone-HMDs z. B. auch im Selbststudium erproben [Sn19]. Problematisch in Bezug auf die Hardware sind jedoch einhergehende **Mobilitätsprobleme** [Da14]. Die aktuell auf dem Markt verfügbare Hardware ist primär für den privaten Gebrauch bestimmt. Folglich sind die Systeme nicht auf den ständigen Transport ausgelegt [JK18]. Eine weitere Herausforderung ist die **eingeschränkte Barrierefreiheit** von Virtual Reality. Formosa et al. stellen (2018) einen signifikanten negativen Einfluss des Alters eines Lernenden und der resultierenden Benutzererfahrung fest. Insbesondere Versuchspersonen mit einem Alter über 35 Jahren hatten Probleme mit der Bedienung und verzeichneten folglich Akzeptanzprobleme [Fo18].

Erschwerend für eine realistische Darstellung von Simulationen ist der **niedrige Reifegrad der Technik**. Insbesondere die realitätsnahe Visualisierung virtueller Patienten in Form von Avataren ist derzeit noch verbesserungswürdig, sodass sich nonverbale Kommunikation in der virtuellen Realität bisher nur begrenzt darstellen lässt. Bspw. für virtuelle Rollenspiele ist dies dennoch ein wichtiger Aspekt [Kl14]. Zusätzlich besteht die Notwendigkeit einer haptischen Erweiterung von VR-Systemen, um den Realitätsgrad zu erhöhen [E6], [Mc18].

5.3 Wirtschaftliche Dimension

Die in den vergangenen Jahrzehnten **gesunkenen Investitionskosten** für VR-Hardware stellen eine Chance für die Nutzung von VR aus Ausbildungsmedium dar. Während die Kosten für ein VR-System im Jahr 2004 noch ca. 45.000 US-Dollar betrugen, waren für die Anschaffung eines gleichwertigen Systems im Jahr 2014 nur noch 1.300 US-Dollar notwendig [JK18]. Durch den Einsatz von VR kann weiterführend die kostenintensive und aus ethischer Sicht kritisch betrachtete **Nutzung von Leichnamen und tierischen Kadavern reduziert** werden [E4]. Ein Medizinstudent verbringt durchschnittlich drei Wochenstunden im Labor, wo er einen Leichnam mit ca. zehn weiteren Studenten teilen muss [Mo17]. Aus der Sicht eines Medizinstudenten sind die Übungen an den Leichna-

men und tierischen Kadavern eine wichtige Praxiseinheit [E5]. Grundlegend besteht jedoch eine ungleichmäßige Verteilung der Leichname, sodass nicht alle Institute über Körperspender verfügen. Wird der praktische Unterricht wie Nähen o. Ä. mit VR erweitert, sinkt der Bedarf an Leichnamen und eine gleichmäßige Verteilung könnte angestrebt werden. Für Universitäten, die weiterhin keine Körperspender erhalten, kann VR als gänzliche Alternative dienen [E3].

Derzeitig angewandte Lehrmethoden wie Rollenspiele sind darüber hinaus mit hohen Personalaufwendungen und -kosten verbunden [Kl14]. Die virtuelle Realität hingegen ermöglicht eine wiederholende Ausführung von Aufgaben mit geringeren Kosten, was zu einer **Steigerung der Kosteneffizienz** führt [Lu13]. VR-Anwendungen entlasten außerdem das Lehrpersonal, wodurch Kapazitäten für die Betreuung der VR geschaffen werden [E1]. Besonders im Vergleich zu traditionellen Simulatoren, wie sie bspw. zum Training von Operationen genutzt werden, sind HMDs deutlich kostengünstiger. Lopes et al. (2016) sprechen in diesem Kontext von „Low-Cost"-Simulatoren [Lo16].

Dem gegenüber stehen jedoch **erhöhte Aufwendungen für Schulungen** des Lehrpersonals und **Wartungen** der **Endgeräte**. Zur Sicherstellung der hygienischen Nutzung von HMDs ist eine regelmäßige Reinigung erforderlich [DSG18]. Derzeitig **fehlende, schulungsbasierte Softwareangebote** auf dem Markt stellen eine besondere Herausforderung dar [Gu16]. Bisher verfügbare Angebote adressieren insbesondere Selbstlerner [E6]. Die Einsatzmöglichkeiten von VR sind daher begrenzt [JK18]. Eine Ursache hierfür liegt in den **hohen Implementierungs**- und **Evaluationsaufwendungen**. Der Gesundheitswesen stellt hohe Anforderungen an seine Lehrmethoden und fordert fundierte Evaluationen, bevor es zu einer Nutzung kommt [Sa18], [E6].

5.4 Politische Dimension

Auf politischer Ebene werden **fehlende staatlich geförderte Projekte** zur Synthese von VR-basierten Schulungen kritisiert. Eine umfassende Untersuchung der Eignung von VR als Bildungsmedium im Gesundheitswesen mit einer staatlichen Beteiligung wird gefordert. Vergangene Verbundprojekte in diesem Bereich wurden insbesondere von Universitätskliniken und Unternehmen aus der privaten Wirtschaft initiiert [E6].

6 Diskussion und Implikationen

Aus soziokultureller Sicht kann mit dem Einsatz von VR als Bildungsmedium im Gesundheitswesen von einer erhöhten Patientensicherheit profitiert werden. Gleichzeitig eignet sich das Medium, um komplexe Zusammenhänge und Vorgänge zu vermitteln. Hierfür fehlt es bisher an entsprechend aufbereiteter Software. Folglich implizieren die Ergebnisse des vorliegenden Beitrages, dass die Zusammenarbeit von VR-Entwicklern,

Akteuren des Gesundheitswesens und der Forschung weiter gefördert werden sollte, damit neue Lehrangebote bereitgestellt werden [E6]. In diesem Kontext kommt dem Staat aufgrund der hohen Implementierungs- und Evaluationsaufwendungen von VR-basierten Aus- und Weiterbildungsangeboten eine tragende Rolle zu. Die Förderung von Forschungsinitiativen im Bereich der immersiven Bildung des Gesundheitswesens kann zur Diffusion der Technologie zu Bildungszwecken beitragen. Das in den Interviews hervorgegangene, große Interesse an VR aufseiten der Lernenden unterstreicht diese Annahme.

Um den dargelegten gesundheitlichen Risiken entgegenzuwirken, ist die Optimierung und Weiterentwicklung bestehender Hardware unumgänglich. Zusätzlich müssen haptische Erweiterungen der Systeme zur Steigerung des Realitätsgrades von virtuellen Lernumgebungen zunehmend berücksichtigt werden. Vor allem die Übung handwerklicher Fähigkeiten, wie sie bspw. im fortgeschritten Medizinstudium oder bei der Weiterbildung von Chirurgen vermittelt werden, erfordert diese Erweiterungen. Die Nutzung virtueller Lernumgebungen benötigt zusätzlich eine einhergehende Betreuungskompetenz aufseiten der Lehrenden. Um die Nutzung von VR als Bildungsmedium zu ermöglichen, sollten zukünftig ausreichend Schulungen zur Vermittlung dieser fehlenden Kenntnisse angeboten werden.

Anzumerken ist, dass VR als alleiniges Bildungsmedium im Gesundheitswesen keinesfalls ausreichend ist. Vielmehr wird „mit VR eine Brücke zwischen der theoretischen Aneignung von Wissen durch das Lesen von Büchern und der ersten Berufspraxis geschlagen“ [E7]. In den Interviews wurden neben der Anatomie insbesondere die Bereiche Physiotherapie, Psychotherapie, Chirurgie sowie sämtliche Fachbereiche mit minimal invasiven Eingriffen wie arthro- und laparoskopische Operationen (Gelenk- und Bauchspiegelung) und 3D-Analyse von CT und MRT als potentielle Anwendungsbereiche der VR identifiziert [E1], [E2], [E7], [E8]. Es bietet sich an, die konkreten Fallszenarien, die in VR abgebildet werden können, in weiteren Untersuchungen zu erheben und auf die Anforderungen eines klinischen Anwendungsfalls zu prüfen. Von dieser Untersuchung können auch privatwirtschaftliche Unternehmen profitieren, die sich auf die Implementierung VR-basierter Lernumgebungen im Gesundheitswesen spezialisieren.

7 Zusammenfassung und Limitationen

Im vorliegenden Beitrag wurden die Chancen und Herausforderungen im Zusammenhang mit der Nutzung von VR als Schulungsmedium untersucht. Hierzu wurde auf die Ergebnisse einer strukturierten Literaturrecherche und einer qualitativen Untersuchung in Form von Experteninterviews zurückgegriffen, die unter Lernenden im Gesundheitswesen durchgeführt wurden. Die Ergebnisse der Literaturrecherche wurden validiert und um die Resultate der Experteninterviews ergänzt.

Die Resultate dieser Studie weisen darauf hin, dass sich die Nutzung von HMDs und VR als Bildungsmedium derzeit am Anfang steht. Dennoch werden mit ihr viele Chancen

assoziiert, zu denen aus soziokultureller Sicht vor allem die Möglichkeit zur Erhöhung der Patientensicherheit sowie eine Verbesserung der Anschaulichkeit von Lerninhalten und daraus resultierende Potentiale in der Lerneffektivität und -effizienz gehören. Technisch bedingt ermöglicht VR eine flexible Darstellung verschiedener Ausbildungsinhalte sowie die Chance, medizinisch-bezogene Kompetenzen auch außerhalb des Krankenhauses zu vertiefen. Die niedrigen Preise für VR-Hardware können zur Diffusion von VR beitragen. Dabei kann der Technologieeinsatz dazu verhelfen, die Kosteneffizienz von Bildungsmaßnahmen im Gesundheitswesen zu steigern. Auch die Bedarfe für schwer zugängliche, kostenintensive Leichname sowie tierische Kadaver können durch die Nutzung von VR reduziert oder ergänzt werden.

Diesen Potentialen stehen jedoch Herausforderungen gegenüber, zu denen fehlende Betreuungs- und Wartungskompetenzen aufseiten der Lehrenden stehen. Problematisch sind ferner die gesundheitlichen Risiken wie die Simulationskrankheit, mit der die Nutzung von VR assoziiert wird. Aus technologischer Sicht stellen die Mobilitätsprobleme der Hardware sowie die eingeschränkte Barrierefreiheit und der derzeitig niedrige Reifegrad der Technik weitere Herausforderungen dar. Des Weiteren hindern aktuell fehlende Softwareangebote sowie notwendige Schulungen für das Lehrpersonal eine weitreichende Implementierung von VR als Bildungsmedium.

Die Resultate unserer Untersuchungen basieren auf einem Literaturreview sowie auf zuvor durchgeführten Experteninterviews. In Bezug auf die Literaturquellen ist anzumerken, dass im ersten Suchschritt fünf Literaturdatenbaken zur Identifikation relevanter Quellen verwendet wurden. Auf diesem Wege wird nicht sichergestellt, dass tatsächlich jegliche für unser Thema relevante Literatur berücksichtigt wurde. Ferner spiegeln die Ergebnisse der Interviews u. a. die subjektiven Auffassungen der Befragten wider. Da die Befragten zum Zeitpunkt der Interviews über wenig Erfahrung mit VR-Systemen verfügten, empfiehlt es sich, die Befragung mit einer größerem Stichprobe fortzuführen und außerdem auf weitere, quantitative Erhebungen mit erfahrenen VR-Anwendern zurückzugreifen. Weiterhin könnte in Interviews mit Startups im VR-Bereich die aktuelle Marktlage fokussiert werden, um den wirtschaftlichen Stand aus Unternehmensperspektive zu betrachten.

Literaturverzeichnis

[Al18] Alfalah, S. F. M.: Perceptions toward adopting virtual reality as a teaching aid in information technology. Education and Information Technologies 23, S. 2633-2653, 2018.

[Bu18] Bucher, K.; Blome, T.; Rudolph, S.; von Mammen, S.: VReanimate II: training first aid and reanimation in virtual reality. Journal of Computers in Education 1, S. 53-78, 2018.

[Bu15] Buń, P.; Górski, F.; Wichniarek, R.; Kuczko, W.; Zawadzki, P.: Immersive Educational Simulation of Medical Ultrasound Examination. Procedia Computer Science 75, S. 186-194, 2015.

[Co12] Consorti, F.; Mancuso, R.; Nocioni, M.; Piccolo, A.: Efficacy of virtual patients in medical education: a meta-analysis of randomized studies. In: Computers & Education 59 (3), S. 1001-1008, 2012.

[Da14] David, O.; Russotto, F.-X.; Da Silva Simoes, M.; Measson, Y.: Collision avoidance, virtual guides and advanced supervisory control teleoperation techniques for high-tech construction: framework design. Automation in Construction 44, S. 63-72, 2014.

[DSG18] Dyer, E.; Swartzlander, B. J.; Gugliucci, M. R.: Using virtual reality in medical education to teach empathy. Journal of the Medical Library Association 106 (4), S. 498-500, 2018.

[FO15] Freina, L.; Ott, M.: A Literature Review on Immersive Virtual Reality in Education: State Of The Art and Perspectives. In: The International Scientific Conference eLearning and Software for Education 1, S. 133-141, 2015.

[Fo18] Formosa, N. J.; Morrison, B. W.; Hill, G.; Stone, D.: Testing the efficacy of a virtual reality-based simulation in enhancing users' knowledge, attitudes, and empathy relating to psychosis. In: Australian Journal of Psychology 70 (1), S. 57-65, 2018.

[GL10] Gläser, J.; Laudel, G.: Experteninterviews und qualitative Inhaltsanalyse, 2. Auflage. Verlag für Sozialwissenschaften, Wiesbaden, 2010.

[Go16] Goldman Sachs: Virtual & Augmented Reality. Equity Research 1, 2016.

[Gu15] Guze, P. A.: Using Technology to Meet the Challenges of Medical Education. Transactions of the American Clinical and Climatological Association 126, S. 260-270, 2015.

[Gu16] Guimaraes, M. de P.; Colombo Dias, D.; Martins, V. F.; Brega, J. R.; Trevelin, L. C.: Immersive and Interactive Simulator to Support Educational Teaching. In: International Conference on Computational Science and Its Applications Part IV, S. 150-160, 2016.

[HP16] Hackett, M.; Proctor, M.: Three-Dimensional Display Technologies for Anatomical Education: A Literature Review. In: Journal of Science Education and Technology 25 (4), S. 641-654, 2016.

[Hu17] Huber, T.; Paschold, M.; Hansen, C.; Wunderling, T.; Lang, H.; Kneist, W.: New dimensions in surgical training: immersive virtual reality laparoscopic simulation exhilarates surgical staff. Surgical Endoscopy 31 (11), S. 4472-4477, 2017.

[Is18] Iserson, K. V: Ethics of Virtual Reality in Medical Education and Licensure. In: Cambridge Quarterly of Healthcare Ethics 27 (2), S. 326-332, 2018.

[Ja16] Janßen, D.; Tummel, C.; Richert, A.; Isenhardt, I.: Towards Measuring User Experience, Activation and Task Performance in Immersive Virtual Learning Environments for Students. In: International Conference on Immersive Learning, S. 45-58, 2017.

[JK18] Jensen, L.; Konradsen, F.: A review of the use of virtual reality head-mounted displays in education and training. Education and Information Technologies 23 (4), S. 1515-1529, 2018.

[Ki13] Kiang, C.; Sundaraj, K.; Sulaiman, M. N.: Virtual reality simulator for phacoemulsification cataract surgery education and training. In: Procedia Computer Science 18, S. 742-748, 2013.

[Kl14] Kleven, N. F.; Prasolova-Førland, E.; Fominykh, M.; Hansen, A.; Rasmussen, G.; Sagberg, L. M.; Lindseth, F.: Virtual operating room for collaborative training of surgical nurses. In: Lecture Notes in Computer Science 8658, S. 223-238, 2014.

[Lo16] Lopes, A.; Harger, A.; Breyer, F.; Kelner, J.: A Natural Interaction VR Environment for Surgical Instrumentation Training. In: International Conference of Design, User Experience, and Usability, S. 499-509, 2016.

[Lu13] Luo, H.; Bian, Y. H.; Zhang, L.; Wang, S. Y.: Application of virtual reality technology to the medical experimental instruction-Taking the example of Wrist-Ankle Acupuncture teaching in Chinese Medicine. In: Proceedings of the 8th International Conference on Computer Science and Education 2013, S. 613-617, 2013.

[Ma19] Mallam, S. C.; Salman, N.; Sathiya, K. R.; Jørgen, E.; Veie, S. E.; Anders, E.: Design of Experiment Comparing Users of Virtual Reality Head-Mounted Displays and Desktop Computers. In: Proceedings of the 20th Congress of the International Ergonomics Association 827, S. 247-257, 2019.

[Mo17] Moro, C.; Štromberga, Z.; Raikos, A.; Stirling, A.: The effectiveness of virtual and augmented reality in health sciences and medical anatomy. In: Anatomical Sciences Education 10 (6), S. 549-559, 2017.

[Mc18] McGrath, J. L.; Taekman, J. M.; Mohan, D.; Bond, W. F.; Kman, N.; Crichlow, A.; Dev, P.; Danforth, D. R.: Using Virtual Reality Simulation Environments to Assess Competence for Emergency Medicine Learners. Academic Emergency Medicine 25 (2), S. 186-195, 2017.

[MD16] Makary, M. A.; Daniel, M.: Medical error – the third leading cause of death in the US. BMJ 353, S. 2139-2144, 2016.

[MN11] Mikropoulos, T. A.; Natsis, A.: Educational virtual environments: A ten-year review of empirical research (1999-2009). Computers and Education 56 (3) 3, S. 769-780, 2011.

[Mo17] Moro, C.; Štromberga, Z.; Raikos, A.; Stirling, A.: The effectiveness of virtual and augmented reality in health sciences and medical anatomy. Anatomical Sciences Education 10 (6), S. 549-559, 2017.

[Ne16] New Media Consotium, Horizon Report: 2016 Higher Education Edition, 2016. URL: https://www.mmkh.de/fileadmin/dokumente/Publikationen/2016-nmc-horizon-report-he-DE.pdf, Stand: 16.04.2019.

[Pa06] Pan, Z.; Cheok, A. D.; Yang, H.; Zhu, J.; Shi, J.: Virtual reality and mixed reality for virtual learning environments. Computers and Graphics 30 (1), S. 20-28, 2006.

[Pa96] Pantelidis, V. S.: Suggestions on when to use and when not to use virtual reality in education. In: VR in the Schools 2 (1), S. 18, 1996.

[PN07] Peng, G. C. A.; Nunes, M. B.: Using PEST analysis as a tool for refining and focusing contexts for information systems research. In: 6th European conference on research methodology for business and management studies, Lisbon, Portugal, S. 229-236, 2007.

[Sa18] Samadbeik, M.; Yaaghobi, D.; Bastani, P.; Abhari, S.; Rezaee, R.; Garavand, A.: The Applications of Virtual Reality Technology in Medical Groups Teaching. In: Journal of advances in medical education & professionalism 6 (3), S. 123-129, 2018.

[Si18] Silva, J. N. A.; Southworth, M.; Raptis, C.; Silva, J.: Emerging Applications of Virtual Reality in Cardiovascular Medicine. JACC: Basic to Translational Science 3 (3), S. 420-430, 2018.

[Si14] Sivan, Y.: Overview: Virtual Reality in Medicine. Journal of virtual world research January, S. 403-440, 2014.

[SL17] Stavroulia, K.-E.; Lanitis, A.: On the Potential of Using Virtual Reality for Teacher Education. In: International Conference on Learning and Collaboration Technologies, S. 173-186, 2017.

[Sn19] Snarby, H.; Gåsbakk, T.; Prasolova-Førland, E.; Steinsbekk, A.; Lindseth, F.: Procedural Medical Training in VR in a Smart Virtual University Hospital. Smart Innovation, Systems and Technologies 99, S. 132-141, 2019.

[So17] Sorathia, K.; Sharma, K.; Bhowmick, S.; Kamidi, P.: A Mobile Based Virtual Reality (VR) Platform to Train and Educate Community Health Workers. In: IFIP Conference on Human-Computer Interaction 2017, Springer, Cham, S. 459-463, 2017.

[St17] Statista, Anzahl der Beschäftigten im Gesundheitswesen in Deutschland in den Jahren 2000 bis 2017. URL: https://de.statista.com/statistik/daten/studie/151723/umfrage/beschaeftigte-im-gesundheitswesen-seit-2000/, Stand: 16.04.2019.

[SL17] Stavroulia, K.-E.; Lanitis, A.: On the Potential of Using Virtual Reality for Teacher Education. In: International Conference on Learning and Collaboration Technologies, S. 199-215, 2017.

[Vo09] Vom Brocke, J.; Simons, A.; Niehaves, B.; Riemer, K.; Plattfaut, R.; Cleven, A.: Reconstructing the giant: on the importance of rigour in documenting the literature search process. European Coference On Information Systems, S. 2206-2217, 2009.

[Wo18] Wong, M. A. M. E.; Chue, S.; Zary, N.; Jong, M.; Benny, H. W. K.: Clinical instructors' perceptions of virtual reality in health professionals' cardiopulmonary resuscitation education. SAGE Open Medicine 6, S. 1-8, 2018.

[ZC19] Zipp, S. A.; Craig, S. D.: The impact of a user's biases on interactions with virtual humans and learning during virtual emergency management training. Educational Technology Research and Development, S. 1-20, 2019.

Zu alt für Informatik?: Seniorinnen und Senioren erobern die digitale Welt

Interesse, Nutzung und Verständnis von Informatiksystemen

Svenja Noichl[1], Ulrik Schroeder [2]

Abstract: Menschen ab 50 Jahren nutzen heutzutage Computer am häufigsten für die Verwendung von Office Produkten, Smartphones werden hauptsächlich zur Kommunikation verwendet und Tablets zur Informationssuche im Internet. Damit unterscheidet sich die Nutzungsweise nicht sehr von der zu Zeiten vor Smartphones und Tablets. Gleichzeitig werden diese Funktionalitäten nicht nur genutzt, sondern es besteht auch ein Interesse daran zu verstehen, wie diese funktionieren. Neben Kommunikationsmöglichkeiten und der Funktionsweise des Internets ist auch Datenschutz und Datensicherheit ein Thema, an dem ein großes Interesse besteht. In unterschiedlicher Ausprägung lassen sich in diesen Nutzungsweisen und Interessen die drei Perspektiven der Dagstuhl-Erklärung, die für die Bildung in der digitalen vernetzen Welt von Bedeutung sind, wiederfinden. Basierend auf diesen Ergebnissen kann ein Konzept zur Vermittlung von ausgewählten informatischen Grundkonzepten erstellt werden, welches an den Alltag der Zielgruppe und ihre Interessen anknüpft und alle drei Perspektiven berücksichtigt.

Keywords: Seniorinnen und Senioren, digitale Bildung, Nutzungsweise von Informatiksystemen, Interesse an Informatik

1 Einleitung

Seniorinnen und Senioren sind längst Teil der digitalen Welt. Ob sie es wissen oder nicht, sie nutzen alle tagtäglich Informatiksysteme. Während in Deutschland bereits 88 % der 50-64-jährigen und 41 % der ab 65-jährigen Smartphones nutzen [Bi17] und somit bewusst Informatiksysteme einsetzen, kommen die anderen unbewusst mit Informatik in Kontakt. Ampeln, Automaten aller Art, z. B. Fahrkarten-, Bank- und Parkscheinautomaten oder digitale Werbeanzeigen in der Stadt, hinter all dem steckt Informatik. Gleichzeitig bieten Informatiksysteme, wie vor allem Smartphones und Tablets, (neue) Möglichkeiten mit Familie und Freunden in Kontakt zu bleiben. Dies wird von vielen Seniorinnen und Senioren in Grundlagen-Workshops als Motivation genannt, sich mehr mit der digitalen Welt und den digitalen Medien auseinanderzusetzen. Auch in gesellschaftlichen Bereichen stehen

[1] RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, Deutschland noichl@informatik.rwth-aachen.de
[2] RWTH Aachen, Lehr- und Forschungsgebiet Informatik 9, Ahornstraße 55, 52074 Aachen, Deutschland schroeder@informatik.rwth-aachen.de

doi:10.18420/inf2019_82

Inhalte der Informatik auf der Tagesordnung. Themen wie Datenschutzgrundverordnung, Urheberrecht, Social Media, usw. sind aus den Nachrichten nicht mehr wegzudenken.
Nach der Dagstuhl-Erklärung sind für die Bildung in der digitalen vernetzten Welt drei Perspektiven von Bedeutung. Diese Perspektiven sind die technologische Perspektive, welche sich damit beschäftigt, wie etwas funktioniert, die anwendungsbezogene Perspektive, die sich damit beschäftigt, wie etwas genutzt wird, und die gesellschaftlich-kulturelle Perspektive, die sich mit der Wirkung auf die Gesellschaft befasst. In diesem Beitrag wird gezeigt, welche der drei Perspektiven bereits im Alltag von Seniorinnen und Senioren auftreten, und welchen weiterer Platz eingeräumt werden sollte. [Br16]
Zu diesem Zweck werden in Kapitel 2 Studien zur Nutzung von digitalen Medien durch Seniorinnen und Senioren vorgestellt. Zum Vergleich mit anderen Zielgruppen werden ebenfalls Ergebnisse aus den KIM- und JIM-Studien dargestellt. Kapitel 3 beschreibt anschließend das Studiendesign und den Fragebogen zur Ermittlung aktueller Nutzungsweisen von Computern, Smartphones und Tablets sowie den Interessen im Bereich Informatik von Seniorinnen und Senioren. Kapitel 4 stellt die Ergebnisse der Befragung im Bereich der Nutzung vor und vergleicht diese mit vorherigen Ergebnissen. Anschließend wird in Kapitel 5 diskutiert, was Seniorinnen und Senioren über Informatik wissen möchten. In Kapitel 6 wird betrachtet, welche der Perspektiven der Dagstuhl-Erklärung sich in diesen Ergebnissen wiederspiegeln und wie ein Konzept zur Vermittlung von Informatikkenntnissen an Seniorinnen und Senioren, welches alle drei Perspektiven berücksichtigt, aussehen kann. Kapitel sieben schließt mit einer Zusammenfassung und einem Ausblick ab.

2 Related Work

Es gibt bereits einige Studien, die sich damit beschäftigen, wozu ältere Erwachsene Geräte wie Computer, Mobiltelefone, Smartphones oder Tablets nutzen. In einer Studie von Selwyn et al. aus dem Jahr 2003 gaben jeweils 27 von 79 Befragten an ihren Computer sehr oft zum Schreiben bzw. Editieren von Briefen oder anderen Dokumenten, bzw. zum Senden und Lesen von E-Mails zu verwenden. [Se03]
Niamh et al. befragten 2012 insgesamt 237 Personen, unter anderem zur ihrer Nutzung von Mobiltelefonen. An dieser Studie nahmen 61 Personen zwischen 50 und 64 Jahren und 48 Personen über 65 Jahren teil. Bei den 50- bis 64-jährigen gaben über 90 % der Befragten an, die Anruf- und Text-Funktion des Mobiltelefons zu verwenden. Unter den über 65-jährigen waren es über 80 % bei der Anruf-Funktion und über 60 % bei der Text-Funktion. Als dritt häufigstes wurde in beiden Gruppen die Kamera genannt. Die Nutzung des Internets fällt in beiden Gruppen mit unter 20 % gering aus. [Ca12]
Mohadisdudis und Ali stellten 2014 ihre Ergebnisse zu einer Studie mit 21 Personen ab 60 Jahren bezüglich deren Handy bzw. Smartphone-Nutzung vor. Unterschieden wurde in dieser Studie zwischen drei unterschiedlichen Gerätetypen („Feature Phone“, „Multimedia Phone“, „Smartphone“). In allen drei Kategorien werden Funktionalitäten zur Kommunikation (hier: Telefonie und SMS) am häufigsten verwenden, gefolgt von der Kamera-Funktion bzw. Fotogalerie. [MA14]

Im Rahmen des österreichischen Forschungsprojekts mobi.senior.A wurden 27 Seniorinnen und Senioren befragt, welche Funktionalitäten ihnen bei Smartphones oder Tablets am wichtigsten sind. Hierzu wurde ihnen eine Auswahl von 36 Funktionen zur Auswahl gegeben. Werden die dort genannten Funktionalitäten in Oberkategorien eingeteilt, zeigt sich hier, Kamera / Fotos als wichtigste Funktionalität. Gefolgt von Wetter, Kalender (Termine, Wecker, . . .), Kommunikation (SMS und E-Mail) sowie Informationssuche im Internet. [Er14]
Ebenfalls im Jahr 2014 führte die Markt- und Meinungsforschungsinstitut GfK im Auftrag von A1 eine Studie zur Mediennutzung und Alltagseinsatz von Smartphones mit Personen über 60 Jahren durch. Den Angaben zu Folge sind SMS verschicken und die Uhr mit 38 % bzw. 37 % die am häufigsten genutzten Funktionalitäten. Von 25 % der Befragten wird das Gerät zur Informationssuche verwendet, 21 % nutzen die Kommunikation mittels E-Mail. [A119]

Studie von	**Jahr**	**Untersuchte Geräte**	**Häufig genutzte Funktionalitäten**
Selwyn et al.	2003	Computer	1. Office Produkte 2. Kommunikation (E-Mail)
Niamh et al.	2012	Mobiltelefon	1. Kommunikation (Anruf, SMS) 2. Kamera
Mohadisdudis und Ali	2014	Mobiltelefon, Smartphone	1. Kommunikation (Anruf, SMS) 2. Kamera
mobi.senior.A	2014	Smartphone, Tablet	1. Kamera 2. Wetter 3. Kalender (Termine, Wecker) 4. Kommunikation (SMS, E-Mail) 5. Informationssuche im Internet
GfK i. A. von A1	2014	Smartphones	1. Kommunikation (SMS, E-Mail) 2. Kalender (Uhr) 3. Informationssuche im Internet

Tab. 1: Am häufigsten genutzte Funktionalitäten von Geräten in den Studien von Selwyn et al., Niamh et al., Mohadisdudis und Ali, mobi.senior.A und GfK

Bei Betrachtung der in Tabelle 1 zusammengefassten und Kategorisierten Ergebnisse kristallisiert sich bereits heraus, dass Kommunikation in unterschiedlichen Formen ein wichtiges Thema ist. Während Computer häufig für die Anwendung von Office Programmen

verwendet werden, zeichnen sich die kleineren und handlicheren Geräte wie Mobiltelefone und Smartphones durch eine häufige Verwendung als Kamera aus. Gerade in den neueren Studien, in denen es hauptsächlich um Smartphones und Tablets geht, gewinnt auch das Thema Informationssuche im Internet zunehmend an Bedeutung. Bezüglich des Themas Kommunikation fällt einzig die mobi.senior.A Studie anders aus als die übrigen Studien zu mobilen Geräten, da hier Kommunikation erst an vierter Stelle genannt wird und nicht an erster.
Nach dieser Betrachtung der Nutzungsweisen von Informatiksystemen durch Seniorinnen und Senioren ist es durchaus interessant zu vergleichen, ob und inwieweit sich diese von der Nutzungsweise von Kindern und Jugendlichen unterscheidet. Hierfür betrachten wir die Ergebnisse der KIM- und JIM-Studie.
In der KIM-Studie 2016 (Kindheit, Internet, Medien), einer Basisstudie zum Medienumgang von Kindern im Alter zwischen 6 und 13 Jahren in Deutschland wurden unter anderem die Handy- und Smartphone-Nutzung, sowie die Tablet-Nutzung im Alltag untersucht. Handys und Smartphones werden in dieser Altersgruppe vor allem zur Kommunikation verwendet. Im Vordergrund stehen dabei vor allem das Telefonieren und Nachrichten austauschen mit den Eltern. 73 % der befragten Kinder nutzen das Gerät auch als Kamera, 68 % spielen Spiele und 47 % nutzen das Internet. Tablets werden am häufigsten zum ansehen von Bildern und Videos verwendet, gefolgt von Spielen und der Internetnutzung. [FLL16]
Bei der JIM-Studie 2018 (Jugend, Information, Medien) handelt es sich um eine Basisuntersuchung zum Medienumgang, bei der Jugendliche im Alter von 12 bis 19 Jahren im Fokus stehen. Eine detaillierte Auflistung nach meistgenutzten Funktionalitäten von Smartphones und Tablets ist hier zwar nicht zu finden, allerdings geht aus der Studie deutlich hervor, dass Kommunikation, insbesondere über WhatsApp und ähnliche Dienste eine bedeutende Rolle in dieser Altersgruppe spielt. [Sü18]
Diese Studien zeigen, dass es zwar kleinere Unterschiede in der Nutzung verschiedener Gerät und innerhalb unterschiedlicher Altersgruppen gibt, gleichzeitig haben sie alle eine Gemeinsamkeit. Unabhängig von Alter oder Gerät ist die Nutzung als Kommunikationsmedium ein wichtiger Aspekt.

3 Studiendesign und Durchführung

Zur Überprüfung der Ergebnisse der in Kapitel 2 vorgestellten Studien zur Nutzung von Computern, Smartphones und Tablets durch Seniorinnen und Senioren und zur zusätzlichen Ermittlung des Interesses dieser Zielgruppe an unterschiedlichen Themen der Informatik, wurde ein Fragebogen erstellt. Um auch die Personen erreichen zu können, die keine Geräte, wie Computer, Smartphone oder Tablet, zur Verfügung haben, bzw. noch nicht geübt im Umgang mit diesen Geräten sind, wurde der Fragebogen in Papierform verteilt. An der Umfrage nahmen 123 Personen ab 50 Jahren teil. Neben demografischen Fragen zum Alter, der Schulbildung sowie dem Beruf stand dabei die aktuelle Nutzung von Informatiksystemen im Vordergrund. Hierzu wurde jeweils zum Computer, Smartphone und Tablet abgefragt, ob, wie häufig und zu welchem Zweck dieses Gerät genutzt wird. Die Teilnehmerinnen und

Teilnehmer wurden hierbei dazu aufgefordert möglichst präzise und detailliert diejenigen Aufgaben und Programme bzw. Apps aufzulisten, welche sie an den entsprechenden Geräten verwenden. Bei der Auswertung wurden diese Angaben in entsprechende Oberkategorien einsortiert um eine Vergleichbarkeit der Antworten zu ermöglichen. Zur Ermittlung des Interesses wurden basierend auf Themen der Informatik, die in der Schule behandelt werden, vgl. z. B. die Bildungsstandards der GI [Be17; Ge08; Rö16], 52 Fragen erstellt. Dabei wurde zum einen darauf geachtet, dass ein Thema von mindestens zwei Fragen abgedeckt wird. Zum anderen war es wichtig bei den Formulierungen darauf zu achten, dass kein besonderes Fachwissen zum Verständnis notwendig ist. Hierzu wurden auch Beispiele verwendet. Eine exemplarische Frage aus dem Fragebogen lautet: Ich interessiere mich für die Unterschiede von Kommunikationsmöglichkeiten (z. B. Unterschied zwischen Telefonat und E-Mail). Zur Beantwortung dieser Fragen stand den Teilnehmerinnen und Teilnehmern eine 6-stufige Likert-Skala zur Verfügung. Dabei stand „1“ für kein Interesse und „6“ für sehr großes Interesse. Um die Möglichkeit abfangen zu können, dass eine Person eine Frage nicht richtig versteht, oder ihr Interesse nicht entsprechend einordnen kann, wurde zusätzlich ein „?“-Feld bereitgestellt. Die Hauptfragestellungen, die mit dieser Umfrage beantwortet werden sollten waren:

1. Wie und wozu nutzen Personen ab 50 Jahren Computer, Smartphones und Tablets?
2. Hat sich die heutige Nutzungsweise im Laufe der Zeit (verglichen zu den in Kapitel 2 betrachteten Studien) verändert?
3. An welchen Themen der Informatik sind Personen ab 50 Jahren (besonders) interessiert?

4 Wie Menschen ab 50 Informatiksysteme nutzen

Wie in Kapitel 2 bereits erläutert existieren Studien dazu, wie Seniorinnen und Senioren Handys bzw. Smartphones verwenden. Allerdings hat sich die Technik sowie der Funktionsumfang von Smartphones und Tablets nach Durchführung dieser Studien weiterentwickelt. Interessant ist es daher noch einmal zu betrachten, inwieweit diese Neuerungen einen Einfluss auf das Nutzungsverhalten von Seniorinnen und Senioren haben. Bei einem Vergleich der drei betrachteten Informatiksysteme Computer, Smartphone und Tablet fällt auf den ersten Blick auf, dass es durchaus Unterschiede in der Nutzung dieser Geräte gibt.
Computer werden am häufigsten für die Arbeit mit Office Produkten verwendet. Von den 76,42 % der Befragten (94 Personen), die Angaben schon einmal einen Computer benutzt zu haben, gaben 62,77 % an, mindestens ein Office Programm zu verwenden. Dabei handelt es sich insbesondere um Programme zum Erstellen von Texten, Tabellen und Präsentationen. 59,57 % der Personen nutzen den Computer zur Kommunikation. Neben zahlreichen Allgemeinen Nennungen wie „Kommunikation“, „Korrespondenz“ oder „Kontakt zu Personen“ sind an dieser Stelle E-Mails hervorzuheben. Auch Briefe wurden von mehreren Teilnehmerinnen und Teilnehmern genannt. Hier ist jedoch nicht ganz klar, ob sie damit

ebenfalls E-Mails meinen, oder die Briefe am Computer verfasst und dann als solche mit der Post versenden. Eine Person gab auch die Nutzung von Skype an. Am dritthäufigsten wird der Computer zum Surfen im Internet verwendet, 47,87 % der Personen gaben dies an. Der Fokus liegt hier auf der Informationssuche. Diese drei Nutzungsgebiete, Office Programme, Kommunikation und Informationssuche im Internet wurden unter den Computernutzern mit großem Abstand am häufigsten angegeben. Die nachfolgenden Tätigkeiten, Online-Banking und Online-Shopping, wurden nur von 18 % bzw. 17 % der Personen angegeben.
Im Bereich der Smartphones ist Kommunikation die mit Abstand am häufigsten angegebene Tätigkeit. Insgesamt gaben 81,30 % (100 Personen) der Befragten an, ein Smartphone zu besitzen. 96 % von ihnen machten zusätzliche Angaben dazu, wozu sie das Gerät verwenden. Allein Kommunikationswege wie E-Mail, SMS und Messenger-Dienste wie WhatsApp, Telegram oder Threema, wurden von 86,46 % dieser 96 Personen genannt. Zusätzlich mit den Personen, die nur den ‚klassischen' Kommunikationsweg der Telefonie angaben, nutzen 95,83 % der Teilnehmerinnen und Teilnehmern ihr Smartphone, um damit zu Kommunizieren. Als zweit häufigstes wurde die Internetnutzung angegeben. Auch hier beziehen sich diese Nennungen auf die Informationssuche. 39,58 % Personen gaben an ihr Smartphone für diesen Zweck zu verwenden. Die Besonderheit des Smartphones, welches durch seine geringe Größe problemlos überall mit hingenommen werden kann, spiegelt sich im Platz drei der Smartphone-Nutzung wieder. Von 32,29 % Personen wird das Smartphone als Kamera und Fotospeicher verwendet. Dies ist der größte hier festzustellende Unterschied in der Nutzungsweise der drei Geräte. Mit jeweils 23 (23,96 %), 20 (20,83 %) bzw. 18 (18,75 %) Nennungen stehen die Funktionalitäten Kalender, Wetter und Navigation beim Smartphone auf den weiteren Plätzen.
Etwa die Hälfte der Befragten, nämlich 60 Personen, gaben an ein Tablet zu besitzen. Weitere Angaben tätigten 85 % von diesen. Auch bei den Tablets sind Kommunikation und Informationssuche im Internet die mit Abstand am häufigsten genutzten Funktionalitäten. Allerdings wird im Gegensatz zum Smartphone das Tablet häufiger zur Informationssuche (32 Nennungen, 58,18 % der Personen) als zur Kommunikation (27 Nennungen, 49,09 % der Personen) verwendet. Am dritthäufigsten werden den Angaben zu Folge Tablets verwendet um sich Nachrichten anzusehen. Damit sind zum Beispiel Tagesschau, NTV oder Online Zeitungen gemeint.
Die nachfolgende Tabelle 2 zeigt noch einmal die drei am häufigsten verwendeten Funktionalitäten von Computern, Smartphones und Tablets im Überblick. Dabei ist jeweils in Klammern die Anzahl an Nennungen angegeben. Die Zahl in den Klammern hinter dem Gerät gibt an, wie viele Prozent der Teilnehmerinnen und Teilnehmer Angaben zu ihrer Nutzungsweise der Geräte machten. Die Zahl in den Klammern hinter den Nutzungsweisen gibt an wie viele Prozent derer, die Angaben zum jeweiligen Gerät getätigt haben, diese Nutzungsweise angaben. Alle Prozentangaben in der Tabelle wurden gerundet.
Vergleichen wir diese Ergebnisse mit den in Kapitel 2 vorgestellten Nutzungsstudien für die Zielgruppe der Seniorinnen und Senioren, können wir die dortigen Ergebnisse bestätigen. Die Weiterentwicklung der Technik in den vergangenen Jahren scheint keinen großen Einfluss auf die Nutzungsweise gehabt zu haben. Computer werden, wie bereits 2003 ermittelt, nach wie vor am meisten für den Einsatz von Office Produkten und zur Kommunikation

	Computer (76 %)	**Smartphone (78 %)**	**Tablet (41 %)**
1	Office Produkte (63 %)	Kommunikation (96 %)	Informationssuche im Internet (63 %)
2	Kommunikation (60 %)	Informationssuche im Internet (40 %)	Kommunikation (53 %)
3	Informationssuche im Internet (48 %)	Fotos und Kamera (32 %)	Nachrichten (30 %)

Tab. 2: Top drei der genutzten Funktionalitäten von Computern, Smartphones und Tablets mit Angabe wie viel Prozent der Personen, die entsprechende Angaben machten, diese Funktionalität nannten

verwendet. Smartphones werden wie in den meisten vorherigen Studien hauptsächlich zur Kommunikation verwendet. Allerdings ist hier die Informationssuche im Internet deutlich relevanter geworden als noch in den vorherigen Studien. Innerhalb des Bereichs der Kommunikation können allerdings Anzeichen der fortschreitenden Digitalisierung erkannt werden. Während in den älteren Studien E-Mail, Telefonie und SMS die Kommunikationswege darstellten, ist dies heute vielseitiger. Wenngleich E-Mail und Telefonie nach wie vor häufig verwendet werden, wird die SMS weitestgehend von modernen Messenger-Diensten wie beispielsweise WhatsApp abgelöst. Hinzu kommen neben der klassischen Telefonie auch die Videotelefonie mittels Skype. Die Tatsache, dass hier auch die Informationssuche im Internet deutlich an Bedeutung gewonnen hat, kann sich darauf zurückführen lassen, dass in den vergangenen Jahren die Möglichkeiten über WLAN sowie günstig und schnell mittels mobiler Datenverbindungen im Internet zu surfen gestiegen sind. Damit unterscheiden sich Menschen ab 50 in ihrem Nutzungsverhalten gar nicht so sehr von Kindern und Jugendlichen. Wie bereits in Kapitel 2 gesehen, nutzen auch diese insbesondere Smartphones überwiegend zur Kommunikation. Der größte Unterschied, der zwischen Kindern und Seniorinnen und Senioren festgestellt werden kann, ist die Nutzung der Geräte zum Spielen. Während Spiele spielen bei Kindern zu den häufigsten Nutzungsweisen, gerade für Tablets, zählt, gaben in der hier vorgestellten Studie gerade einmal 5 von 96 Smartphone-Nutzern (5,21 %) und 12 von 51 Tablet-Nutzern (23,53 %) an das Gerät zum Spielen zu verwenden.

5 Was Menschen ab 50 über Informatik wissen möchten

Wie in Kapitel 3 beschrieben, wurden die Teilnehmerinnen und Teilnehmer in dem Fragebogen dazu befragt, wie hoch ihr Interesse an unterschiedlichen Themen der Informatik ist. Die gestellten Fragen wurden hierzu in 11 Themenschwerpunkte zusammengefasst. Die folgende Auflistung zeigt diese Themenschwerpunkte und stellt kurz vor, was hier darunter zu verstehen ist:

1. Algorithmen: Probleme und Aufgaben können in der digitalen Welt durch vorgegebene Abläufe gelöst werden. Dabei kann es unterschiedliche Abläufe für die Lösung eines Problems geben.

2. Automaten: Hier geht es um das Verständnis davon, wie Automaten, die mittlerweile Einzug in viele Lebensbereiche erhalten haben, wissen, was sie als nächsten tun müssen, d. h. wie sie von einem Zustand in den nächsten Wechseln. Zudem geht es darum die Interaktion mit Automaten verstehen und beschreiben zu können.

3. Darstellung von Daten: Daten können auf unterschiedliche Weise dargestellt werden. Ein gutes Beispiel hierfür sind unterschiedlich grafisch aufbereitete Statistiken. In diesem Themenbereich geht es um die unterschiedlichen Möglichkeiten von Darstellungsweisen, deren Vor- und Nachteile, sowie deren Interpretation.

4. Datenschutz und Datensicherheit: Hier stehen zum einen Verschlüsselungsverfahren, Maßnahmen zum Schutz vor unerwünschtem Zugriff auf persönliche Daten und Seriosität von Onlineressourcen im Vordergrund. Zum anderen geht es auch darum sicher mit personenbezogenen Daten umgehen zu können und Maßnahmen zum unerwünschten Zugriff bewerten zu können.

5. Informatik, Mensch und Gesellschaft: Die Fragen zu diesem Themengebiet beschäftigen sich mit der angemessenen Nutzung von Technologie im Alltag sowie deren Vor- und Nachteilen, dem Einfluss von Technologie auf den Menschen und umgekehrt, sowie dem bewussten Umgang mit Werken anderer, z. B. Urheberrecht von Bildern oder Videos.

6. Informatiksysteme: Hier geht es um die Bestandteile sowie Beschreibung der Funktions- und Arbeitsweise von Informatiksystemen, das Abschätzen der Notwendigkeit ihrer Nutzung und der passenden Softwareauswahl für die zu lösenden Aufgaben sowie das selbstständige erschließen von neuen Informatiksystemen.

7. Internet und Netzwerke: Durch das Internet wird eine weltweite Vernetzung von Computern ermöglicht. So können sich auch Menschen weltweit miteinander vernetzen. Neben diesem Aspekt steht die Funktionsweise des Internets und die Arbeit mit Internetdiensten im Fokus. Auch Informationsbeschaffung aus Onlinequellen und die Erkennung von Werbung fallen in diesen Bereich.

8. Kommunikation mit Informatiksystemen: Hier ist gemeint, dass Personen mithilfe von Informatiksystemen miteinander kommunizieren. Dies passiert beispielsweise durch E-Mails, Telefonate oder WhatsApp Nachrichten. Die unterschiedlichen Kommunikationswege haben dabei zum einen verschiedene Eigenschaften, sie können z. B. synchron oder asynchron sein, zum anderen haben sie jeweils Vor- und Nachteile.

9. Kommunikation über Informatiksysteme: Hier geht es darum sich mit anderen Menschen über Themen der Informatik bzw. über Informatiksysteme unterhalten zu können. Das fundierte Mitreden können in der digitalen Welt ist hier der zentrale Aspekt.

10. Organisation von Daten und Dokumenten: Backups, Replikationen und Versionierung können einem ungewollten Datenverlust vorbeugen. Zudem ist es sinnvoll Dokumente so abzuspeichern, dass sie zu einem späteren Zeitpunkt leicht wiedergefunden werden können.

11. Umgang mit Problemen: Zu Problemen kann es immer kommen, auch beim Umgang mit Informatiksystemen. Häufig zeigen die Geräte in einem solchen Fall Fehlermeldungen an. Hier geht es darum diese Fehlermeldungen interpretieren zu können.

Ein erhöhtes Interesse äußerten die Teilnehmerinnen und Teilnehmer in der Befragung in den Themenschwerpunkten Datenschutz und Datensicherheit, Kommunikation mit Informatiksystemen, Kommunikation über Informatiksysteme, Informatiksysteme, Internet und Netzwerke, Organisation von Daten und Dokumenten sowie Internet und Netzwerke. Dabei bedeutet erhöhtes Interesse hier, dass der Median aller gegebenen Antworten für eine Frage bei mindestens 4 und somit in der oberen Hälfte der Skala liegt. Es gilt jedoch zu beachten, dass nicht alle Aspekte der genannten Themenschwerpunkte ein erhöhtes Interesse aufweisen. Im Bereich Datenschutz und Datensicherheit wurden für die Aspekte des sicheren Umgangs mit personenbezogenen Daten und der Einschätzung der Seriosität von Onlineressourcen Werte von >= 5 erzielt. Allerdings sinkt das Interesse bezogen auf Maßnahmen und Mechanismen für Datenschutz und Datensicherheit auf Werte von 3 bzw. 4. In den Bereichen Kommunikation mit Informatiksystemen und Kommunikation über Informatiksysteme wurde deutlich, dass die aktive Teilhabe an der digitalen Gesellschaft ein wichtiger Aspekt für die Teilnehmerinnen und Teilnehmer ist. Die Aspekte mit anderen in Kontakt zu treten und zu bleiben sowie mitreden zu können erzielte im Median werden von 4 bzw. 5. Für den Bereich Informatiksysteme riefen diejenigen Fragen, in denen es um die richtige Nutzung und Interaktion mit den Geräten geht das höchste Interesse hervor (Median von 4 bzw. 5). Der Median für Fragen im Bereich Internet und Netzwerke lag bezüglich der Erkennung von Werbung sowie der Informationsfindung, –beschaffung und –bewertung im Internet ebenfalls bei 4 bzw. 5. Für den Bereich Organisation von Daten und Dokumenten wurden Werte >= 5 erreicht für Aspekte wie abspeichern und wiederfinden von Daten, Schutz vor versehentlichem Löschen von Dateien sowie unterschiedliche Möglichkeiten zur Datenspeicherung. Auch im Bereich Umgang mit Problemen, in welchem die Interpretation von Fehlermeldungen in der Nutzung von Informatiksystemen im Vordergrund steht, wurde ein Median von 5 erreicht.
Zusätzlich konnte festgestellt werden, dass das Interesse von Personen, die noch nie ein Tablet benutzt haben, bei den folgenden Aspekten signifikant höher war, als bei den, die schon einmal ein Tablet benutzt haben bzw. dieses regelmäßig nutzen. Diese Personen zeigten ein höheres Interesse an Fragen im Bereich Darstellung von Daten. Hier insbesondere bei unterschiedlichen Darstellungsformen für Daten und deren Interpretation, z. B. in Diagrammen und Statistiken. Auch weisen sie ein höheres Interesse an dem Verständnis von Funktions- und Arbeitsweisen von Informatiksystemen sowie des Internets und der Bewertung der Qualität von Informationen aus dem Internet auf.
In diesen Ergebnissen spiegelt sich wieder, dass die Befragten Personen unter anderem

in den Bereichen ein erhöhtes Interesse besitzen, in denen auch die häufig genutzten Funktionalitäten liegen. Die Kommunikation und die Informationssuche im Internet sind jedoch nicht nur für diejenigen Interessant, die bereits Computer, Smartphone oder Tablet benutzen, sondern auch für diejenigen, die dies noch nicht tun. Ein ebenso relevantes Thema ist Datenschutz und Datensicherheit.

6 Dagstuhl und Informatik für Menschen ab 50

Mit den in den Kapiteln 4 und 5 dargestellten Ergebnissen zeigt sich, dass alle drei Bereiche der Dagstuhl-Erklärung bereits im Alltag der Seniorinnen und Senioren Anwendung finden, bzw. dass sie an Aspekten aller drei Bereiche Interesse haben, dabei ist die technologische Perspektive jedoch am geringsten ausgeprägt. Am stärksten spiegelt sich die anwendungsbezogene Perspektive wieder. Diejenigen, die bereits ein oder mehrere Geräte (Computer, Smartphone, Tablet) verwenden, nutzen diese bereits in ihrem Alltag für unterschiedliche Aufgaben. Auch das Interesse daran, wie diese Informatiksysteme genutzt werden ist hoch. Die befragten Personen möchten wissen, wie sie mit den Geräten interagieren können, wie sie diese zur Kommunikation mit anderen Menschen nutzen können oder wie sie das Internet nutzen können, um Informationen zu finden. Die gesellschaftlich-kulturelle Perspektive lässt sich hier im Interesse daran bei informatischen Themen mitreden zu können wiederfinden. Ebenso machen sich die Befragten Gedanken zu Themen wie Datenschutz und Datensicherheit und sie wollen Werbung erkennen. Die Teilhabe an der digitalen Gesellschaft, zum einen durch ein Verständnis der Wirkung von Informatik auf die Gesellschaft, beispielsweise durch personalisierte Werbung aufgrund von persönlichen Daten im Internet, oder durch die Nutzung digitaler Medien zur Kommunikation und zum in Kontakt bleiben mit anderen Menschen, spielen eine wichtige Rolle für Seniorinnen und Senioren. Bezüglich der technologischen Perspektive viel auf, dass insbesondere Personen, die noch nie ein Tablet genutzt haben, ein signifikant höheres Interesse an Aspekten dieser Perspektive zeigten. Bei ihnen gab es z. B. ein erhöhtes Interesse daran die Funktionsweise des Internets zu verstehen.

Im Folgenden wird ein Konzept bestehend aus vier Modulen vorgestellt, in welchem alle drei Perspektiven der Dagstuhl-Erklärung berücksichtigt werden, um Seniorinnen und Senioren Informatikkenntnisse zu vermitteln. Die Module bauen dabei aufeinander auf und haben das Ziel den Wunsch nach Kommunikation aus Sicht der Informatik zu beleuchten. Dabei ist es gerade für Seniorinnen und Senioren, insbesondere für diejenigen, die noch nicht viele Berührungspunkte mit Geräten wie Computern, Smartphones oder Tablets hatten, wichtig einen Bezug zu deren Alltag herzustellen.

Das erste Modul beschäftigt sich mit der Frage 'Wie kann ich mit anderen kommunizieren?' Motiviert wird dieses Modul damit, dass die Seniorinnen und Senioren mit ihrer Familie oder ihren Freunden in Kontakt bleiben möchten. In bereits durchgeführten Tablet-Workshops wurde bei einer kurzen Vorstellungsrunde häufig der Wunsch mit entfernt lebenden Familienmitgliedern oder Freunden in Kontakt bleiben zu können als Grund für die Teilnahme

genannt. In diesem Modul werden unterschiedliche Methoden zum Kommunizieren behandelt und verglichen. Angefangen mit den der Zielgruppe gut bekannten, nicht digitalen Methoden wie Brief und Anruf, werden die Eigenschaften und Unterschiede auf Methoden wie E-Mails, Videotelefonie oder Instant Messenger übertragen. Am konkreten Beispiel des Teilens oder Weiterleitens von Bildern soll zudem für das Thema Urheberrecht sensibilisiert werden.
Das zweite Modul befasst sich darauf aufbauend mit der Frage 'Wie kommen meine Nachrichten zu ihrem Bestimmungsort?' Hierbei geht es insbesondere um diejenigen Kommunikationswege, die das Internet mit einbeziehen. Es sollen die Grundzüge der Funktionsweise des Internets vermittelt werden, sodass ein grundlegendes Verständnis davon erlangt werden kann, wie die Datenübertragung im Internet funktioniert. Gleichzeitig soll herausgestellt werden, dass es mithilfe von Computernetzwerken, am Beispiel des Internets, möglich ist weltweit Computer und so auch Menschen miteinander zu vernetzen.
Das dritte Modul beschäftigt sich mit der Frage 'Wie kann ich mein mobiles Gerät mit dem Internet verbinden, um eine Nachricht verschicken zu können?' Hier geht es darum unterschiedliche Varianten kennenzulernen, um insbesondere mobile Geräte wie Smartphones oder Tablets mit dem Internet zu verbinden. Hier gibt es vor allem die Möglichkeiten WLAN und mobile Daten. Zusätzlich soll es in diesem Modul auch darum gehen, wer wann welches WLAN nutzen kann.
Im vierten Modul geht es dann um das Thema Datenschutz und Datensicherheit. Es befasst sich mit der Frage 'Ist das ganze denn auch sicher?' In den vorangegangenen Modulen wird an unterschiedlichen Stellen deutlich, dass es Accounts gibt oder Passwörter nötig sind. Diese Thematik soll einen ersten Einstig in den Bereich Datenschutz und Datensicherheit liefern. Im ersten Teil des Moduls geht es darum herauszustellen, was Passwörter sind, wozu man diese benötigt und welche Kriterien es zur Erstellung sicherer Passwörter gibt. Im Weiteren wird dann stärker für das Thema Datenschutz und Datensicherheit sensibilisiert. Hierzu wird beispielsweise geklärt, was unter diesen Begriffen zu verstehen ist und welche Konsequenzen es geben kann.
In allen vier Modulen werden Aspekte der technologischen und gesellschaftlich-kulturellen Perspektive aktiv behandelt. Es steht jeweils die Funktionsweise der vorgestellten Aspekte im Vordergrund der Module. Am Ende jedes Moduls wird dabei auf deren Wirkung auf die Gesellschaft eingegangen. Die anwendungsbezogene Perspektive spiegelt sich in der kurzen Beschreibung der Module nicht direkt wieder, ist aber aufgrund der Umsetzung der Module ebenfalls gegeben. Die Module werden in Form von Android Apps umgesetzt. Dies bietet zum einen die Möglichkeit eine Learning Analytics Anbindung zu schaffen, zum anderen können auf diese Weise die Inhalte selbstständig und im eigenen Tempo erarbeitet werden. Gleichzeitig bietet es den Seniorinnen und Senioren die Möglichkeit den Umgang und die Nutzung mobiler Geräte stetig zu trainieren. Mithilfe von Simulationen können ihnen die Funktionsweisen veranschaulicht werden oder sie können unter realen Bedingungen das theoretisch erworbene Wissen anwenden, indem sie z. B. eine E-Mail verschicken oder im Internet nach einer bestimmten Information suchen müssen. Abbildung 1 zeigt die Kompetenzen, die in den vier Modulen vermittelt werden sollen. Bei der weiteren

Ausgestaltung der Module müssen diese in konkrete und überprüfbare Lernziele überführt werden.

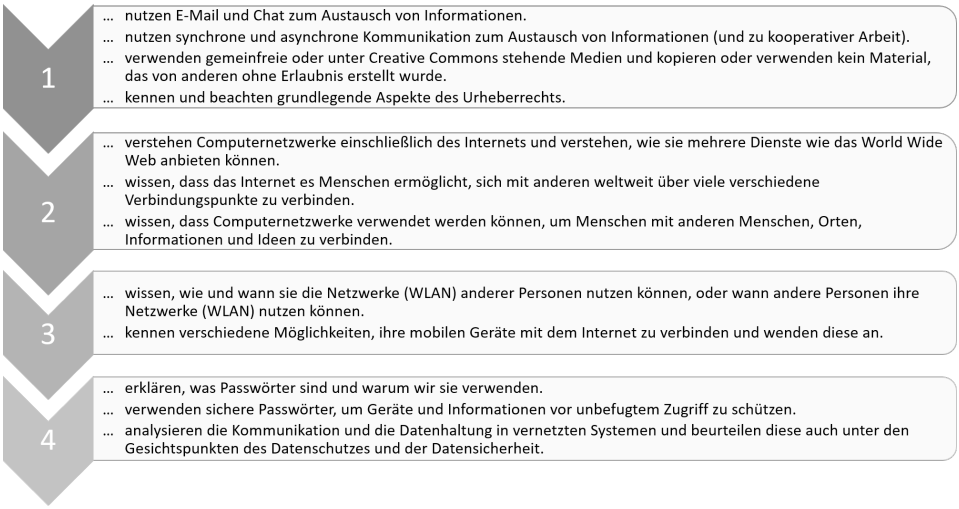


Abb. 1: Kompetenzen der vier Module

7 Zusammenfassung und Ausblick

Mithilfe einer Befragung von 123 Personen ab 50 Jahren konnte festgestellt werden, dass sich das Nutzungsverhalten derjenigen, die Computer, Smartphones oder Tablets verwenden im Vergleich zu ähnlichen Studien im Laufe der Zeit nicht verändert hat. Lediglich im Bereich der Kommunikation mittels dieser Geräte konnte festgestellt werden, dass mit Entwicklung neuer Kommunikationswege, diese zunehmend auch Verwendung finden. Kommunikation und Informationssuche im Internet zählen bei allen drei betrachteten Geräten zu den Top 3 der meist verwendeten Funktionalitäten. Auch bei der Frage nach dem Interesse der Zielgruppe sind diese Aspekte sehr wichtig. Zudem besteht ein großes Interesse an den Themen Datenschutz und Datensicherheit. Anhand der steigenden Zahlen von Personen in dieser Altersgruppe, die sich Smartphones und Tablets anschaffen und der hier vorgestellten Befragung wird deutlich, Seniorinnen und Senioren möchten an der digitalen Welt teilhaben und sind zudem interessiert mehr über Informatik zu erfahren, um die Geräte besser nutzen und verstehen zu können. Diesem Interesse kann mit Kursangeboten, oder wie hier vorgestellt, mit Lern-Applikationen entgegengekommen werden. Das hier vorgestellte Konzept bietet den Vorteil, dass alle drei Perspektiven der Dagstuhl-Erklärung berücksichtigt werden. Gerade durch die Vermittlung der zugrundeliegenden Funktionsweisen und nicht nur der Vermittlung reiner Nutzungskompetenzen, kann so auch eine Übertragbarkeit des gelernten auf andere Informatiksysteme ermöglicht werden.
Im Weiteren wird das hier vorgestellte Konzept inhaltlich weiter ausgestalten und technisch

umgesetzt. In einem iterativen Entwicklungsprozess mit Beteiligung der Zielgruppe soll sowohl die Verständlichkeit der aufbereiteten Inhalte für die Zielgruppe evaluiert und verbessert werden. Zum anderen wird evaluiert, welche technischen Möglichkeiten zur Darstellung dieser Inhalte verwendet werden können und wie eine Learning Analytics Anbindung die Lerner unterstützen kann.

Literatur

[A119] A1: Die Seniorenstudie von A1, 2019, URL: https://newsroom.a1.net/news-die-seniorenstudie-von-a1?id=59351&menueid=12658, Stand: 19. 06. 2019.

[Be17] Best, A. et al.: Kompetenzen für informatische Bildung im Primarbereich. Beilage zu LOG IN 38/(189/190), 2017.

[Bi17] Bitkom: Anteil der Smartphone-Nutzer in Deutschland nach Altersgruppen im Jahr 2017, 2017, URL: https://de.statista.com/statistik/daten/studie/459963/umfrage/anteil-der-smartphone-nutzer-in-deutschland-nach-altersgruppe/, Stand: 19. 06. 2019.

[Br16] Brinda, T.; Diethelm, I.; Gemulla R.and Romeike, R.; Schöning, J.; Schulte, C.: Dagstuhl-Erklärung: Bildung in der digitalen vernetzten Welt. Gesellschaft für Informatik eV, 2016.

[Ca12] Caprani, N.; Doyle, J.; O'Grady, M.; Gurrin, C.; O'Connor, N. E.; Caulfield B.and O'Hare, G. M.: Technology use in everyday life: implications for designing for older users. 2012.

[Er14] Erharter, D.; Jungwirth, B.; Knoll, B.; Schwarz, S.; Posch, P.; Xharo, E.: Smartphones, Tablets, App für Seniorinnen und Senioren. Assistenztechnik für betreutes Wohnen. AAL Testregion Westösterreich. Tagungsband zum uDay XII/, S. 221–235, 2014.

[FLL16] Feierabend, S.; LFK, T. P.; LFK, T. R.: KIM-Studie 2016 Kindheit, Internet, Medien. Basisstudie zum Medienumgang. 2016.

[Ge08] Gesellschaft für Informatik e. V.: Grundsätze und Standards für die Informatik in der Schule. Bildungsstandards Informatik für die Sekundarstufe I. LOG IN 28/(150/151), 2008.

[MA14] Mohadisdudis, H. M.; Ali, N. M.: A study of smartphone usage and barriers among the elderly. 2014 3rd International Conference on User Science and Engineering (I-USEr)/, 2014.

[Rö16] Röhner, G. et al.: Bildungsstandards Informatik für die Sekundarstufe II. Beilage zu LOG IN/(183/184), 2016.

[Se03] Selwyn, N.; Gorard, S.; Furlong, J.; Madden, L.: Older adults' use of information and communications technology in everyday life. Ageing & Society 23/(5), S. 561–582, 2003.

[Sü18] Südwest, M. F. Hrsg.: JIM-Studie 2018. Jugend, Information, Medien. Basisuntersuchung zum Medienumgang 12-bis 19-Jähriger. 2018.

Development of a senior-friendly training concept for imparting media literacy

Sebastian Wilhelm[1] Dietmar Jakob[2] Melanie Dietmeier[3]

Abstract: The use of digital solutions to support rural areas, and in particular elderly people, is the goal of the *'Digitales Dorf'* and *'BLADL'* research projects. This work assessed seniors' media literacy in two model communities; with the result that fear of fraudster and lack of knowledge are the most common causes that prevent elderly people from using digital technologies. Based on these evaluation results, a combined training offer of *tutorial* and *digital consultation hour* was developed and evaluated.

Keywords: Digitization; Media literacy; Senior education

1 Introduction

The current progress of digitization can be recognized in almost all areas of life. As the younger generation grows up with this progress and learns how to use digitalization in schools as well, older people are increasingly having problems using the new technologies.

According to the online study of ARD / ZDF in 2018, there were around 63.3 million Internet users in Germany [AR18]. However, there is a clear trend that the proportion of the users in the age group 60+ drops significantly, confirming the original thesis. *Figure 1* shows the proportion of Internet users in Germany who use the Internet daily or almost daily, by age group in 2018.

However, even seniors could experience relief in everyday life and improve their family and social contacts with digital technologies. As an example, here can be mentioned the communication with children, family and friends via instant messaging.
In fact, there is a strong barrier to the use of digital technologies and participation in courses. The elderly must change proven habits and processes and complement them with digital technologies. For more than two-thirds of people over the age of 60, these habits are an

This work was founded by the *Bavarian State Ministry of Family Affairs, Labour and Social Affairs*

[1] Deggendorf Institute of Technology, Technology Campus Grafenau, Hauptstraße 3, D-94481 Grafenau, sebastian.wilhelm@th-deg.de

[2] Deggendorf Institute of Technology, Technology Campus Grafenau, Hauptstraße 3, D-94481 Grafenau, dietmar.jakob@th-deg.de

[3] Deggendorf Institute of Technology, Technology Campus Grafenau, Hauptstraße 3, D-94481 Grafenau, melanie.dietmeier@th-deg.de

doi:10.18420/inf2019_83

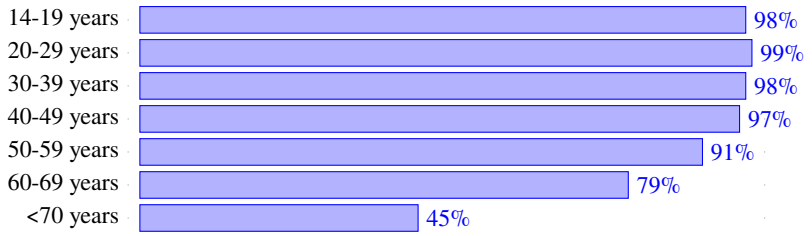


Fig. 1: Proportion of Internet users in Germany using the Internet every day or almost every day, by age group [St18].

important part of their lives and give them security in their daily lives. There is also no professional pressure for seniors to acquire digital knowledge. Therefore, there is a need to reduce ambivalence towards digital media and strengthen digital skills in old age. These include, for example, free WLAN access, the availability of broadband networks (especially in structurally weak regions) and access to terminals in retirement facilities (in nursing homes, volunteer agencies, neighborhood centers, multi-generational homes, senior citizen's offices, etc.) or public places that are also visited by older people, so-called study rooms (Internet cafes, chat rooms) [Kl16].

As part of this work, a training concept has been developed to provide citizens 55+[4] with targeted media latency. *'BLADL'* is linked to the *'Digitales Dorf'*[5] project and therefore focuses mainly on rural regions. The training concepts are tested and evaluated in two communities in the Bavarian Forest, Frauenau[6] and Mauth-Finsterau[7].

In a first step, a market analysis of existing education offers was created (*Section 2*). In a subsequent questionnaire campaign in the communities of Frauenau and Mauth-Finsterau, citizens (55+) were asked about the use of digital technologies, their needs, obstacles and wishes (*Section 3*). Subsequently, further education and support services were developed, which are currently being carried out with the aim of improving the media literacy of seniors (*Section 4*).

[4] No fixed border

[5] Further information about the project on `www.digitales-dorf.bayern`

[6] Municipality in the district of *Regen* / *Lower Bavaria* with a total of 2665 citizens from which 686 persons are in the age of 65 or older [Ba16a]

[7] Municipality in the district of *Freyung-Grafenau* / *Lower Bavaria* with a total of 2279 citizens from which 479 persons are in the age of 65 or older [Ba16b]

2 Analysis of existing education offers

A total of 40 educational offers from 31 different providers where examined in an market analysis of the existing educational offers for seniors[8]. Providers are mainly educational institutions, associations or private companies.

Depending on the offer, the group size of the participants varies between 5 and 15 participants and the offer ranges from one 2-hour session to six 3-hour sessions. The main topics are communication, internet apps, data protection and data security as well as online shopping.

Educational providers use the methods of lectures and exercises, individual consultations, group work, online training, self-study courses or videos.

> Older people prefer a written course booklet, so they can read in peace what they have seen and done under guidance. [Ku18].

Furthermore, the analysis suggests that the following conditions have been taken into account in individual offers in order to make participation for seniors enjoyable and successful:

- short course times (maximum 2-3 hours)
- allow sufficient breaks
- self-contained topics - no modular structure
- enough time for questions and exercises to plan
- learning at the residence of the senior
- small groups with individual care.

3 Survey on the determination of media literacy

At the start of the project in 2018 a total of 681 citizens 55+ of the municipality of Mauth-Finsterau were sent questionnaires by post. Citizens were asked about the use of digital technologies, needs, obstacles and desires to develop tailor-maid support offers.

At the end of the collection period, 145 questionnaires were submitted (response rate 21%). In addition, there are 89 questionnaires of a second champaign in the community Frauenau *Table 1* shows the distribution of the survey participants by age group.

Below, some of the results of this survey will be presented.

[8] The research was done via the internet

age group	amount	percentage
55-64 years	112	≈ 48%
65-74 years	90	≈ 38%
75-84 years	22	≈ 9%
no answer	10	≈ 4%

Tab. 1: Distribution of the survey participants by age group.

3.1 Use of digital technologies

In the first part of the survey, the frequency of using digital technologies such as mobile phones, smartphones, computers (PCs, laptops, notebooks) and tablets was questioned. Multiple selection was possible. *Table 2* shows the detailed results.

frequency	mobile phone	smartphone	computer	tablet
every day	19%	41%	39%	18%
weekly	6%	3%	12%	6%
less common	16%	2%	13%	5%
never	12%	24%	18%	36%
no answer	47%	30%	18%	35%

Tab. 2: Using digital technologies by frequency and device type. Multiple selection was possible.

If the item *never* has been selected for use or if no answer has been given, it can be assumed that there are no corresponding devices. Therefore *Table 3* shows adjusted results.

	mobile phone	smartphone	computer	tablet
regularly	41%	46%	64%	29%
never	59%	54%	36%	71%

Tab. 3: Frequency of using digital technologies (adjusted results). Multiple selection was possible.

It can not be ruled out that the difference between mobile phones and smartphones was clear to every respondent. However, a total of 216 respondents (≈ 92%) indicated that they use at least a mobile phone or a smartphone. Only 6 participants (≈ 3%) indicated that they did not use any of the requested devices or did not answer the question.

3.2 Use of applications and services

The questionnaire also asked which digital applications and services citizens are already using.

Figure 2 shows a list of the services and applications used by citizens, sorted by frequency.

Citizens explained mainly that they would use telephony and messenger / e-mail services with subsequent search and order options on the Internet. Another main application area is

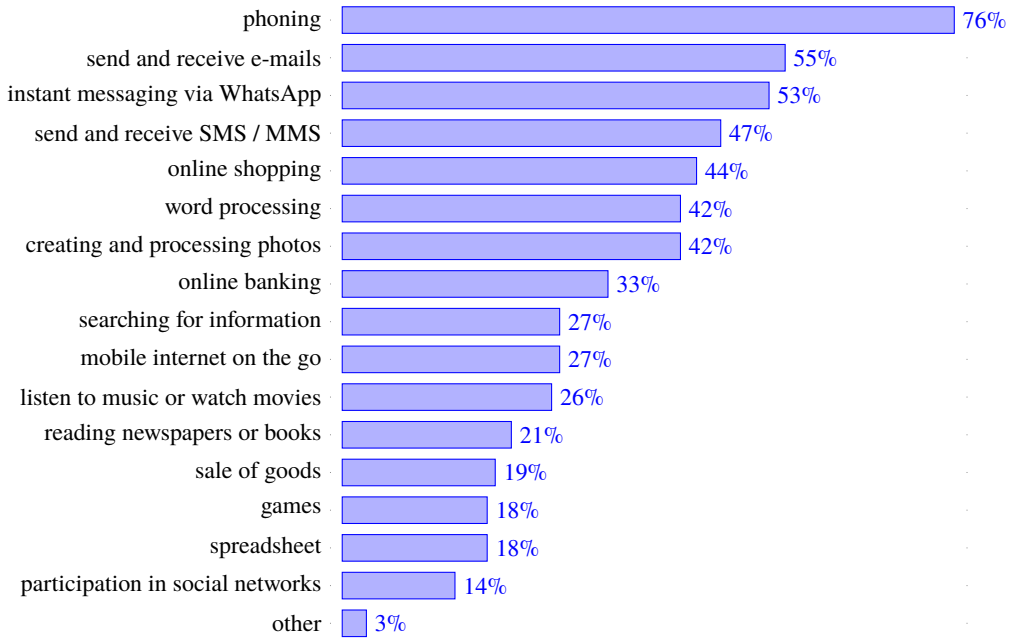


Fig. 2: Survey results for the questions of which digital applications and services are currently been used from the citizens. Multiple selection was possible.

word processing, which is extremely useful on the PC. Photographing, as well as editing and sending photos is one of the main applications.
Rather, fewer respondents use online banking. An explanation for this can be found in *Figure 3*, where nearly half of the respondents stated, that they are afraid of being abused by fraudsters. Games or even services like eBay and Facebook play a minor role.

A comparison with similar studies can only be differentiated based on the various studied age groups. Furthermore, the researched geographic location is also crucial in the assessment. In this regard, the results are only partially comparable.

Looking at other studies when using the applications and services, there are both parallels and contradictions. The study of people over the age of 50, 'Technology use in everyday life: implications for older users (2012)' [Ca19], indicates that 65% of respondents use e-mail services. This study also takes into account the age group of 50-55 year-olds, which was disregarded in the present study. However, the age group is already more tech-oriented due to the use of modern technologies in professional life and explains the discrepancy of the results. When using messaging services (e.g. WhatsApp), the results are almost identical to the surveys in the mobi.Senior.A project [Am15]. Interestingly enough, the usage rate of online banking services, music and video streaming, as well as the creation and editing of photos with the surveys of Caprani [Ca19] differs only marginally. According to a study by DIVSI [SI16], online shopping is used by 41% of people over the age of 60 and is almost identical to the available figures (44%). Of minor importance in the use include services such as reading newspapers, games and social networks (Facebook).

A comparison of the results of different studies in the use of digital services is shown in *Table 4*.

service	DIT	Mobi.senior.A	Telefonica	A1	DIVIS	Caprani
phoning	76%	n/a	n/a	n/a	n/a	90%
send and receive e-mails	55%	77%	78%	21%	87%	65%
instant messaging	53%	55%	28%	n/a	26%	n/a
send and receive SMS	47%	88%	n/a	38%	n/a	n/a
online shopping	44%	n/a	29%	n/a	41%	n/a
creating and processing photos	42%	96%	n/a	n/a	n/a	47%
online banking	33%	n/a	30%	n/a	n/a	32%
searching for information	27%	n/a	41%	21%	86%	63%
listen to music or watch movies	26%	62%	23%	n/a	6%	28%
reading newspapers or books	21%	n/a	50%	n/a	9%	n/a
games	18%	n/a	45%	n/a	9%	n/a
participation in social networks	14%	22%	18%	n/a	n/a	n/a

Tab. 4: Comparison of our (DIT) survey results on services used with the studies [Am15], [Te], [DB14], [SI16], [Am15]

3.3 Reasons preventing the use of digital technologies

An essential reason why seniors do not use digital technologies is the subjective difficulty experienced. In addition, DIVSI [SI16] comes to this conclusion. The mobi.senior.A project [Am15] states that usability issues are hurdles where older people would fail if there is not someone to show them how to proceed.

As part of our survey citizens were also asked for reasons that prevent them from using digital technologies.
The most frequently mentioned was the fear of fraudsters. This result is not cited in any of the studies [Ch17, SI16, Te, Ca19, Am15, DB14]. The results of related studies generally refer to data protection and data security aspects that hinder its use.
Other important reasons for non-users are the lack of knowledge and competence in dealing with media, misdirected on-demand support and complicated handling of the equipment. Slow Internet access or overinvestment are not considered as relevant.

The detailed results are shown in *Figure 3*.

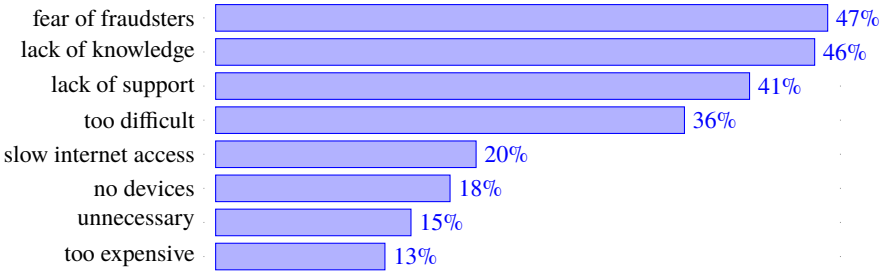


Fig. 3: Survey results for the question of which reasons does the citizens prevent from using digital technologies. Multiple selection was possible.

3.4 Desired additional support

The survey has shown that the majority of digital technology users receive help and support from their own family, such as partners and children. Rather less said to seek help from neighbors or acquaintances on the Internet itself and commercial providers. Even the own problem solving has only a law value. Some do not seek support at all, but simply stop using it.

The results for the question of what service citizens are looking for are shown in *Figure 4*.

Often, support within the family is not enough and additional help is required. The survey shows that citizens prefer exercises with a coach and have the opportunity to attend a digital

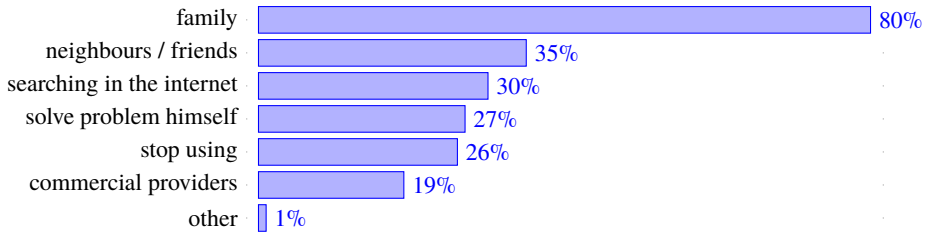


Fig. 4: Survey results for the question of where the participants currently are searching for support around the use of digital technologies.
The question was answered by 195 people. Multiple selection was possible.

consultation hour. There is also a need for traditional offerings such as courses or seminars in educational institutions. The interest on digital worksheets for tablets, PCs or books is rather low. Also video courses seem to be of lesser importance.

The results for the question of what service citizens are looking for are shown in *Figure 5*.

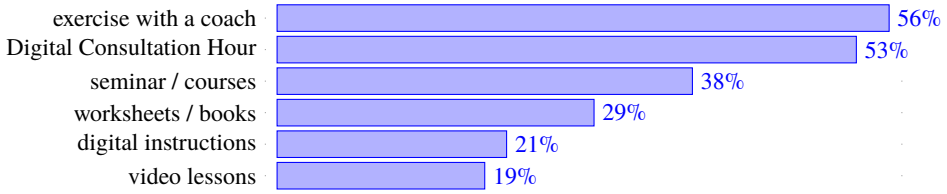


Fig. 5: Survey results for the question of which supporting offers the citizens are searching for.
The question was answered by 111 people. Multiple selection was possible.

The interpretation of the results shows that trainings or seminars on digital topics alone are not enough. Additional support formats must be offered in order to deepen what you have learned or to be able to solve any problems that may arise. The Telefonica study [Te] and mobi.senior.A [Am15] reached the same conclusions. Older people do not just have to have somebody who shows them the most important functions and control options, but also help them relatively quickly, if they do not know how to continue, according to the findings of the project mobi.senior.A [Am15]. This is even more evident in the Telefonica [Te] study. Authors conclude that support services are needed for people who are not attending a course or have no one in their area who can be contacted when seeking help with regard to such issues.

4 Implementation

Taking into account the analysis of existing education offerings (*Section 2*) and the results of the previous citizens‘ survey (*Section 3*), a training concept has been developed that consists of two components. On the one hand, the *tutorials* are held in small groups with a personal trainer and on the other hand, a regular *digital consultation hour* was set up. Particular care was taken to ensure that all offers are suitable for seniors. Among other things, the barrier-free accessibility of the training rooms or the age-appropriate graphic design of the training documents (e.g. font size / contrast) must be ensured.

All offers are free of charge and non-binding to citizens. In order to address people who currently do not have a suitable device, a pool training devices (smartphones, tablets and laptops) has been set up, which can be made available to the participants for the tutorials.

In the following, the two implemented supporting offers are described in more detail.

4.1 Tutorials

The essential knowledge transfer takes place in the context of *tutorials*. Therefore, courses take place in small groups (6-8 participants), where each session deals with a self-contained thematic block[9]. Furthermore, the courses take place locally (in the communities), so that distances are eliminated.

The content of the tutorials focuses on functionality, technology-specific background knowledge (e.g. describing end-to-end encryption) is only treated if it is inevitable necessary. Accordingly, the course topics are always described in a purposeful way and advertised such e.g. ‘short messages from and to children, grandchildren and friends via WhatsApp’. This should overcome the first inhibition threshold for the direct use of new technologies. The participants should be able to draw an immediate added value from the course.

During the tutorial, the participant is introduced step by step to a practical goal (e.g. sending an image via WhatsApp). The participants will receive exercises during the tutorial, that they can complete independently on their own devices and thus have the opportunity to deepen their knowledge directly. For a single course, a period of no more than two hours is regularly provided to promote sustainable knowledge acquisition and not to overstrain the participants.

In order to receive further information or to look up all learning contents more compactly, all course participants receive a course booklet, which contains the content of the training and partly further printed (also technical) information.

The *‘BLADL’* project has offered courses on the following topics:

[9] No hierarchical course structure. Each session can be vised independently to the participation of other courses

- How do smartphones, tablets & co. work?
- Short messages from and to children, grandchildren and friends via WhatsApp
- Short messages from and to children, grandchildren and friends via e-mail
- Information about doctors, medicine and holidays - using the tablet on the Internet
- Basics of word processing with Microsoft Word
- Take beautiful pictures with your smartphone, edit and save
- Don't be afraid of online banking
- Safe shopping on the internet

By conducting the courses, the inhomogeneity of user devices (especially for smartphones and tablets) is a particular challenge. Often, menus, functions or labels are very different, which makes it difficult for the participant to follow. The tutor therefore needs to pay particular attention to teaching the logic of a system and not having fixed step-by-step introductions that may not work on some user devices. Technically, mirroring the screen with a tutorial device on the projector can be very helpful in solving these difficulties. Further, the tutor may need to support some participants during the tutorial on their own devices so that everyone can follow the course. It should be noted, however, that particularly with smartphone and tablet courses, a very high degree of teacher´s affinity (mainly technical) is necessary to assist the participants on their own device when needed.

To participate in the tutorials a registration is required to guarantee the small course size. This opportunity is also used, among other things, to derive some course-relevant information in advance (e.g. is a Google-Account available?) and to prepare the training accordingly.

An initial evaluation[10] of the tutorials shows that the participants receive the concept very well and that almost all participants are very satisfied with the offer. The feedback also shows that the participants are looking for more, and after the first inhibition threshold is overcome, they are also interested in advanced courses.

4.2 Digital consultation hour

In addition to the tutorials, a digital consultation hour was set up. This format complements the tutorials so that citizens have the opportunity to receive further support for small questions and problems around digital technologies.

The *digital consultation hour* within the *'BLADL'* project takes place on a weekly rhythm for two hours per community. Citizens can also take advantage of this offer without registration, giving them the opportunity to receive free assistance from experienced staff.

[10] In March 2019 with currently 305 participants at tutorials

It is interesting to know that the offer is very well received in Frauenau, with an average of 6 people per week, whereas on average in Mauth-Finsterau only 1 person uses the same offer.

The offer should include more detailed information on specifics and individual technical questions and problems. In addition, the feeling of security should be conveyed in order to have a contact person in case of problems with a device. This is to encourage the citizens to experiment with the devices themselves.

The long-term goal of the digital consultation hour is to encourage exchanges between citizens themselves and to encourage seniors to assist each other in technical matters.
In the community Frauenau already first successes were achieved here. Even during the regular digital consultation hour, the seniors help each other with minor concerns.

5 Conclusion and Future Work

As part of the *'BLADL'* research project existing training courses for seniors were evaluated (Section 2) and citizens 55+ in the communities of Frauenau and Mauth-Finsterau were asked about their preferences and wishes for (additional) training on the Internet in the field of digital technologies (Section 3). Based on these results, a two-part training concept was developed, carried out and initially evaluated, consisting of *tutorials* on the one hand and a regular *digital consultation hour* on the other hand (Section 4).

Until March 2019, there were already 37 tutorials with more than 300 participants and 27 digital consultation hours with around 100 participants, which were successfully carried out and initially evaluated. It should be noted, that the two-part concept is very well accepted by the citizens and the course evaluation shows very good results.
The problems reported in the digital consultation hour are also becoming increasingly complex, suggesting that citizens are using for themselves and experimenting with new digital technologies. Citizens are even beginning to communicate the offer to other seniors, so the customer-base for the tutorials and digital consultation hour is continually growing.

The next step is to examine the transferability of the two-part training concept to other regions. For this purpose, the financial feasibility is first examined. An attempt is made to determine to what extent the participants in the tutorial (as well as the digital consultation hour-participants) are prepared to pay for such an offer and how citizens‘ expectations change when they have to pay for the offer. In addition, various public / voluntary approaches (such as the distribution to multi-generation houses or the continuation of an organized neighborhood assistance) need to be assessed.

In parallel, it is planned to conduct a second survey in the municipalities of Frauenau and Mauth-Finsterau to assess the sustainability of the training concept and the transfer of knowledge.

Acknowledgements

This work was founded by the *Bavarian State Ministry of Family Affairs, Labour and Social Affairs*[11] within the project *Digitales Dorf - BLADL / Besser Leben im Alter durch digitale Lösungen*[12].

References

[Am15] Amann-Hechenberger, Barbara; Buchegger, Barbara; Erharter, Dorothea; Felmer, Viktoria; Fitz, Bernadette; Jungwirth, Bernhard; Kettinger, Marlene; Schwarz, Sonja; Knoll, Bente; Schwaninger, Teresa; Xharo, Elka: Tablet & Smartphone: Seniorinnen und Senioren in der mobilen digitalen Welt. Technical report, Österreichisches Institut für angewandte Telekommunikation (ÖIAT), Büro für nachhaltige Kompetenz B-NK GmbH, ZIMD – Zentrum für Interaktion, Medien und soziale Diversität, 2015.

[AR18] ARD, ZDF: , Anzahl der Internetnutzer in Deutschland in den Jahren 1997 bis 2018 (in Millionen). Statista, 2018. Accessed: 2019-03-13.

[Ba16a] Bayerisches Landesamt für Statistik: Demographie-Spiegel für Bayern. Berechnungen für die Gemeinde Frauenau. Technical report, Bayerisches Landesamt für Statistik, 2016.

[Ba16b] Bayerisches Landesamt für Statistik: Demographie-Spiegel für Bayern. Berechnungen für die Gemeinde Mauth. Technical report, Bayerisches Landesamt für Statistik, 2016.

[Ca19] Caprani, Niamh; Doyle, Julie; O'Grady, Michael; Gurrin, Cathal; O'Connor, Noel; Caufield, Brian; O'Hare, Gregory: Technology Use in Everyday Life: Implications for Designing for Older Users. 06 2019.

[Ch17] Christine, Weiß; Julian, Stubbe; Catherine, Naujoks; Sebastian, Weide: Digitalisierung für mehr Optionen und Teilhabe im Alter. Technical report, Bertelsmann Stiftung, 2017.

[DB14] Dandrea-Böhm, Livia: , Die Seniorenstudie von A1, 2014.

[Kl16] Klein, Ludger: Runder Tisch „Aktives Altern – Übergänge gestalten". Fachgespräch am 21. Januar 2016 in Frankfurt am Main: Digitale Kompetenz älterer Menschen - Dokumentation, 2016.

[Ku18] Kubicek, Herbert: Leitfaden - Digitale Kompetenzen für ältere Menschen. Technical report, Telefónica Deutschland Holding AG, Stiftung Digitale Chancen, 2018.

[SI16] SINUS-Instituts Heidelberg: DIVSI Ü60-Studie Die digitalen Lebenswelten der über 60-Jährigen in Deutschland. Technical report, Deutsches Institut für Vertrauen und Sicherheit im Internet, 2016.

[St18] Statistisches Bundesamt: . Private Haushalte in der Informationsgesellschaft – Nutzung von Informations- und Kommunikationstechnologien, 2018.

[Te] Telefónica Deutschland Holding AG, Stiftung Digitale Chancen: Digital mobil im Alter. Technical report, Telefónica Deutschland Holding AG, Stiftung Digitale Chancen. Survey only on usage of tablets.

[11] www.stmas.bayern.de

[12] www.digitales-dorf.bayern

Extended Abstracts

Technikbegleitung. Aufbau von Initiativen zur Stärkung der Teilhabe Älterer im Quartier

Elisabeth Bubolz-Lutz[1], Janina Stiel[2]

Abstract: Dieser Beitrag möchte anhand eines 4-jährigen FuE-Projekts auf die Notwendigkeit für und die Möglichkeiten von digitaler Bildung im Alter aufmerksam machen. Digitale Bildung wird im Allgemeinen als Aufgabe von Schulen oder beruflicher Aus- und Weiterbildung betrachtet – also für Kinder und Erwachsene. Häufig übersehen wird dabei, dass ältere Menschen die Mehrheit der Offliner in Deutschland stellen. Moderne Technologien haben das Potenzial für mehr Lebensqualität im Alter und können Teilhabe ermöglichen. Aufgezeigt wird, wie digitale Bildung im Alter gelingen kann.

Keywords: Alter(n); digitale Bildung; Teilhabe; freiwilliges Engagement; Technikbotschafter/innen

Ältere Menschen machen sich zunehmend mit der Internetnutzung, IKT und anderen Technologien vertraut. Dennoch gehörten auch im Jahr 2018 noch 40 Prozent der ab 60-Jährigen zu den „Offlinern“ [In19] – das entspricht in etwa 9,3 Millionen Menschen [St]. Dabei ist digitale Teilhabe – also das Beteiligtsein an der Nutzung des Internets, digitaler Medien und moderner Technologien – ungleich verteilt. Zu den „Offlinern“ zählen eher Frauen , Hochaltrige, Ältere mit geringer formaler Bildung, Alleinlebende und Menschen mit gesundheitlichen Einschränkungen [TRWW16]. Häufigste Gründe gegen die Nutzung sind eine fehlende Nutzenwahrnehmung, die Einschätzung, dass die Geräte zu komplex und kompliziert zu bedienen sind und damit der Lernaufwand hoch wäre sowie Bedenken gegenüber der Datensicherheit und dem Datenschutz [De16]. Den ungleichen digitalen Teilhabechancen zu begegnen, ist von Bedeutung, um Menschen nicht weiterhin von gesellschaftlichen Entwicklungen zu exkludieren, wenn Wahlmöglichkeiten zwischen digital und analog stetig abnehmen (Haushaltsgerätesteuerung, elektr. Steuererklärung, Schließen von Bank- und Fahrkartenschaltern). Eine falsche Annahme ist es zudem, dieses „Problem“ werde sich von allein lösen. Künftige Generationen älterer Menschen werden zwar eine höhere Technik- und Medienkompetenz aufweisen als vorige, aber „die technische Entwicklung wird weiter ständig voranschreiten und Fortschritte werden mit immer größerer Geschwindigkeit aufeinander folgen. [...] So wird die Notwendigkeit, auch vorhandene digitale Kompetenzen stets zu erweitern und anzupassen, eher größer als kleiner“ [Fo16].

Während es auf der einen Seite Strategien braucht, tatsächlich für die Lebenswelt Älterer relevante Technologien partizipativ zu entwickeln, gilt es auf der anderen Seite Strategien der digitalen Bildung für Ältere zu konzipieren und zu erproben, die es besonders auch

[1] Forschungsinstitut Geragogik, Spichernstr. 18a, 40476 Düsseldorf, bubolz-lutz@fogera.de
[2] Bundesarbeitsgemeinschaft der Senioren-Organisationen, Servicestelle Digitalisierung und Bildung für ältere Menschen, Thomas-Mann-Str. 2-4, 53222 Bonn, stiel@bagso.de

doi:10.18420/inf2019_84

technikdistanten Älteren ermöglichen, einen Einstieg in die digitale Welt zu finden. Denn wie Ältere lernen (wollen) verändert sich gegenüber früheren Lebensphasen und bedarf spezifischer Vorgehensweisen [Bu10]. Das Handbuch [BLS18] präsentiert das Ergebnis des Teilprojekts „Technikbegleitung“ des inter- und transdisziplinären BMBF-Projekts „QuartiersNETZ". Es zeigt, wie in Städten und Kreisen Initiativen aus freiwillig Engagierten auf- und ausgebaut werden können, die „Technikbegleitung“ anbieten: Speziell qualifizierte „Technikbotschafter/innen“ zeigen Älteren in ihrem Wohnumfeld, wie sich technische Geräte und digitale Medien handhaben lassen – in Einsteigerkursen, Sprechstunden oder einer 1:1 Begleitung, bei Bedarf auch zu Hause. Die Technikbotschafter/innen selbst sind technisch versierte ältere Menschen und damit zugleich Rollenvorbilder. Dargestellt wird, wie sich Technikbegleitung in ein kommunales Gesamtkonzept integrieren lässt, wie Freiwillige zu Technikbotschafter/innen qualifiziert werden können, wie eine Lernplattform zur Weiterqualifizierung und Organisation der Initiative genutzt werden kann, welche Aufgaben in der Praxis entstehen (Öffentlichkeitsarbeit, Angebotsspektrum festlegen, Lernorte ausstatten, Kooperationen schließen, Evaluation) und wie kognitive und motivationale Veränderungen im Alter beim Technik-Lernen berücksichtigt werden können.

Deutschlandweit gibt es bereits über 300, zumeist selbstorganisiert entstandene Initiativen verschiedenen Professionalisierungsgrades und sie sind bisher das Erfolgsmodell zur Förderung von Technik- und Medienkompetenz im Alter. Anders als Kurse traditioneller Bildungsanbieter (z.B. VHS) erreichen die Initiativen auch technikdistante Gruppen, die sich von „moderner Technik“ eher überfordert fühlen. Die Initiativen benötigen jedoch eine nachhaltige Anbindung an kommunale Strukturen, niedrigschwellige Lernorte, finanzielle Förderung, sowie eine landes- und bundesweite Vernetzung. Im Ergebnis kann Technikbegleitung ein Baustein einer notwendigen bundesweiten Bildungsstrategie für die digitale Bildung Älterer sein.

Literaturverzeichnis

[BLS18] Bubolz-Lutz, Elisabeth; Stiel, Janina: Technikbegleitung: Aufbau von Initiativen zur Stärkung der Teilhabe Älterer im Quartier. Dortmund, 2018.

[Bu10] Bubolz-Lutz, Elisabeth; Gösken, Eva; Kricheldorff, Cornelia; Schramek, Renate: Geragogik: Bildung und Lernen im Prozess des Alterns. Kohlhammer, Stuttgart, 2010.

[De16] Deutsches Institut für Vertrauen und Sicherheit im Internet: DIVSI Ü60-Studie: Die digitalen Lebenswelten der über 60-Jährigen in Deutschland. Hamburg, 2016.

[Fo16] Forschungsgesellschaft für Gerontologie e.V.: Abschlussbericht zur Vorstudie „Weiterbildung zur Stärkung digitaler Kompetenz älterer Menschen". Dortmund, 2016.

[In19] Initiative D21: D21 Digital Index 2018/2019. Berlin, 2019.

[St] Statistisches Bundesamt, https://service.destatis.de/bevoelkerungspyramide, 18.06.2019.

[TRWW16] Tesch-Römer, Clemens; Weber, Constanze; Webel, Henry: Nutzung des Internets durch Menschen in der zweiten Lebenshälfte. Berlin, 2016.

Autorenverzeichnis