Autor

Wolfgang Gerhard Köhler

Titel

Challenges of Efficient and Compliant Data Processing:

Assuring legal access to data

Kumulative Dissertation

zur Erlangung des Doktorgrades (doctor rerum politicarum)

an der Wirtschafts- und Sozialwissenschaftlichen Fakultät

der Universität Potsdam

Datum der Disputation

13.02.2024

Betreuer: Prof. Dr. habil. Christoph Rasche (Universität Potsdam)
Prof. Dr. Christian Schultz (Technische Hochschule Wildau)

Gutachter: Prof. Dr.-Ing. habil. Norbert Gronau (Universität Potsdam)
Prof. Dr. Uta Herbst (Universität Potsdam)
Prof. Dr. Eric Kearney (Universität Potsdam)

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-62784
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-627843

Acknowledgments

Many dear people have accompanied me through the last four years and thereby supported my writing of this thesis.

First, I would like to thank Prof. Dr. Christoph Rasche and Prof. Dr. Christian Schultz for their excellent supervision and great support and their valuable feedback on my research at every stage of completing the entire thesis. I would like to thank Prof. Dr.-Ing. habil. Norbert Gronau, Prof. Dr. Uta Herbst and Prof. Dr. Eric Kearney for reviewing this thesis.

My gratitude belongs to Andreas Herzig, Dr. Ljuba Kerschhofer-Wallner, and all my colleagues at Deloitte for their advice during my studies and for giving me the flexibility and freedom to advance my research.

Finally, I want to thank my dear friends Aaron, Alexander, and Patrick, who have shaped me and my thoughts in countless discussions.

To my wonderful family and my parents, I thank you for all your support at any time and over any distance. You have always encouraged, supported, and strengthened me immeasurably.

Table of Contents

List of Figures

List of Tables

Part A – Research Summary

1. Introduction

Problem Statement

The volume of collected data and data triangulation has increased dramatically. The digitalization of functions and auxiliary services primarily drives product innovations in the automotive environment. The latest technologies allow private companies and public authorities to use data for unprecedented purposes in their operations. The process of digitalization is fundamentally changing the business and the private world, with implications for people and companies everywhere in the world and in every industry. Digital transformation is about integrating smart data into everything we do. Rich data plays a crucial role in data-driven economies of scale (Karnouskos, Kerschbaum, 2018). Since a large volume as well as high quality of data help companies to improve products and services innovative data usage raises the value for the customer and monetization opportunities for companies. With the perpetual cycle of channeling data into information to learn continuously the data-driven world will never be offline.

More and more technologies can be used to generate, archive and process digitized information. A large number of scientific articles have been reporting for years on the potential benefits for organizations of processing large volumes of data, these include creating value and meeting desired customer needs (De Luca, Herhausen, Troilo, & Rossi, 2020; Del Vecchio, Mele, Passiante, Vrontis, & Fanuli, 2020), and acquiring competitive advantage (e.g., Libert, 2013; Manyika et al.).

In the data realm, the processing of personal data as the most valuable data category plays a significant role. Technologies capable of generating, archiving, and processing digital information about people and their daily lives are constantly evolving and becoming more sophisticated. The "datafication" of people refers to processes that use data from people's online interactions (van Dijck, 2014). Using personal data to improve health, increase safety, improve productivity, or enhance the quality and individuality of products and services are commonly described examples of technological development. Our virtual and physical behavior is tracked more actively and extensively today than ever before. Rarely does this happen in the interest of the end-user, even if they accept the risks voluntarily (Phelan C., Lampe C., and Resnick P., 2016). Ethical, moral, and societal norms are being challenged in the context of personal data processing and management. Related topics are, for example, the legitimacy and legality of the processing of personal data, the right to privacy, and the informational self-determination of data subjects. Concerns have been brought to the forefront of political debate. Not only by

famous cases of widespread misuse of personal data, such as the Cambridge Analytica scandal in 2018, in which Facebook was implicated when a data profiling company used information from the platform to influence voting and other behaviors (Isaak, Hanna, 2018). Some high-profile academic publications (e.g., Zuboff, 2019) also outline images of the "death of privacy" or ever-refining profiles on the internet, delineating a "wicked" social future (Tutton, 2017) in contrast to papers and studies on technological progress and its benefits.

Since increased attention does not necessarily lead to better transparency of current or ongoing processing operations (Spiekermann, Acquisti, Böhme, 2015), lawmakers are trying to keep up with technological advances, and data privacy legislation is developing around the world concerning the fundamental rights of individuals to protect personal data. The regulations aim at greater transparency and integrity and at strengthening the position of those concerned in the enforcement of their rights, above all the right to privacy regarding the processing of personally identifiable information (PII), often referred to as personal data. However, translating ethical guidelines and policies into regulatory mechanisms and standards remains challenging (Mittelstadt, Allo, Taddeo, Wachter, and Floridi, 2016).

On the one hand, companies find themselves in the challenging situation of processing, storing, and managing personal data according to data privacy laws, and, in the case of multinational companies, they must adhere to multiple country-specific rules. To lay the foundation for this, organizations must implement various privacy laws quickly and comprehensively. Furthermore, sometimes different and changing requirements of the relevant markets increase the degree of complexity. However, uninterrupted end-to-end access to legally usable data is already indispensable in many areas today. On the other hand, there are the interests of the users as natural persons with the right to privacy, expecting to exercise real control over their data at any time. Users are not unjustifiably claiming the right to decide how their data is used and benefit from advantages achieved in the case of personal data processing.

For the further development of legal frameworks and guidance for a successful implementation of data privacy requirements in practice, more knowledge is needed about the discrepancy between the desired and the actual regulatory impact. In addition to the effective implementation of privacy laws and to ensure ongoing access to as much high-quality data as possible, insights about influencing factors regarding users' willingness to share their data are needed. In a nutshell, data management must meet three conditions: First, data usage must be legal and compliant. Second, data usage must be legitimate from the viewpoint of ethics. Third, data usage must be value-generating from a business perspective.

Research Objectives

The circumstances described above led to the four objectives of this thesis, which have been examined through seven papers and one book chapter.

Research Overview

RO= Research Objective; RQ = Research Question

RO 1: To better understand the challenges of implementing privacy legislation.

RQ 1: What are the major challenges & unsolved problems in implementing data privacy laws?

RQ 2: What are the possible solutions to improve the implementation of data *privacy requirements*?

- Köhler, W., Schultz, C., Rasche, C. (2020). Das 100% Problem im Datenschutz. *24. Interdisziplinäre Jahreskonferenz zu Entrepreneurship, Innovation und Mittelstand (G-Forum)*, Karlsruhe
- Gümüs, C., Köhler, W., Schultz, C.; Rasche, C. (2021). *Konzept eines CRISP-DM-Modells zur ganzheitlichen Datenschutzbetrachtung unter Anwendung von Data Mining*. In: Reussner, R. H., Koziolek, A. & Heinrich, R. (Hrsg.), INFORMATIK 2021. Gesellschaft für Informatik, Bonn. (S. 1021-1035). ISBN 978-3-88579-708-1.
- Köhler, W., Schultz, C., Rasche, C. (2024, forthcoming). *Data Governance Insights – The Magic Triangle of Data Governance*, Deloitte GmbH
- Koehler, W. (2023). *Developing Digital Products with Compliance-Driven Personal Data Integration.* In: Bitran, I., Bitetti, L., Conn, S., Fishburn, J., Huizing, E., Ritala, P., Torkkeli, M., Yang, J., (Eds.) Proceedings of XXXIV ISPIM Innovation Conference, Ljubljana, Slovenia, 04 June 2023. ISBN 978-952-65069-3-7.

RO 2: To better understand the factors influencing customers' willingness to share personal data.

RQ 1: What factors influence customers' willingness to share personal data?

RQ 2: How can these factors help to ensure and expand legal access to personal data?

- Köhler, W., Schultz, C., Rasche, C. (2020). *When Handing out Presents is not Enough! – Influencing Factors on the User´s Willingness to Share Data for Connected Car Services*. *24*. Interdisziplinäre Jahreskonferenz zu Entrepreneurship, Innovation und Mittelstand (G-Forum), Karlsruhe

- Köhler, W., Schultz, C., Rasche, C. (2024, forthcoming). *Access to Legal Data as a Dynamic Capability - The Case of Connected Car Services. Industry and Innovation.* (submitted)

RO 3: To better understand the role of data privacy for digital entrepreneurship.

RQ: What business opportunities arise for entrepreneurs from the development of privacy legislation?

- Koehler, W., Schultz, C., & Rasche, C. (2022). *Data are the fuel for digital entrepreneurship - but what about data privacy?*. In: Keyhani, M.,Kollmann, T., Ashjari, A., Sorgner, A., Hull, C. (eds.) Handbook of Digital Entrepreneurship (pp. 306-322). Edward Elgar Publishing. https://doi.org/10.4337/9781800

RO 4: To better understand the importance of a multidisciplinary examination of data regulation.

RQ 1: How has the significance of GDPR evolved within the research landscapes of law, economics, and computer science, and what insights can be gleaned from this evolution?

- Köhler, W. (2023, forthcoming/ working paper). *Mapping the Field Across Disciplines: Data Protection Research in Law, Economics and IT.*

In relation to the inaugural research objective - enhancing comprehension of the primary challenges in enforcing data privacy laws and suggesting appropriate technical solutions for advancing the execution of data protection regulations - additional research is requisite. Thus, this dissertation delves into the paramount operational obstacles engendered by the persistent evolution of global privacy legislation to deriv guidelines to bolster organizations' adherence to privacy standards.

Paper 1, entitled “Das 100% Problem im Datenschutz,” emphasizes the transparency and documentation obligations ensuing from the European General Data Protection Regulation (EU GDPR), which was instituted on May 25, 2018. This study deals with the inherent complications in the pragmatic implementation of documentation requisites as complete compliance seems nearly unattainable for companies.

Paper 2, entitled “Konzept eines CRISP-DM-Modells zur ganzheitlichen Datenschutzbetrachtung unter Anwendung von Data Mining,” delves into the prospective

application of data mining technology to support and optimize the execution of documentation obligations - in this instance, the provisions of Article 30, GDPR.

The whitepaper, named “The Magic Triangle of Data Governance” (featured in Deloitte Data Governance Insights), explicates additional prerequisites "outside the privacy box" essential for achieving data governance, and new questions in this context. To answer those is particularly relevant for globally operating organizations.

These trio of papers are complemented by the research paper (Paper 4), “Developing Digital Products with Compliance-Driven Personal Data Integration.” It outlines pragmatic strategies for the integration of personal data into digital products. By embracing transparent and ethical data practices, a balance between the advantages of data collection and privacy apprehensions can be established. A predictive adoption of certain practices allows organizations to devise user-focused, innovative digital products and services that cater to user needs while safeguarding privacy.

These studies collectively enhance the understanding on the quality and efficiency of implementation projects and the continual adherence to privacy norms. They further elucidate technical opportunities to support data privacy governance and compliant data usage. These peer-reviewed conference papers were presented at the Interdisziplinäre Jahreskonferenz Entrepreneurship, Innovation und Mittelstand (G-Forum) in Karlsruhe (virtually) on October 2, 2020 (Paper 1), and at the Jahreskonferenz der Gesellschaft für Informatik 2021 (INFORMATIK 2021), Panel: Recht und Technik - Datenschutz im Diskurs (RuT2021) in Berlin (online) on September 27, 2021 (Paper 2). Paper 3 is slated for publication by Deloitte in the final quarter of 2023. The fourth research paper was presented at the XXXIV ISPIM Innovation Conference on June 04, 2023.

Since more data trumps better algorithms (Rajesh et al., 2012; Recchia and Jones, 2009) and the development of privacy legislation worldwide shifts control over personal data more and more into customer’s hands, a better understanding of what factors concern users regarding the processing of their data and drive their decisions to disclose personal data is necessary. Concerning the second research objective, this dissertation shows in a modified macro model on a large data sample in the context of CCS, which factors influence the willingness of individuals to share their data with companies. This, in turn, shows companies what specific management measures can help them to access as much personal data as possible to improve digital products and strengthen their position in the market. The peer-reviewed conference paper “When Handing out Presents is not Enough! - Influencing Factors on the User´s

Willingness to Share Data for Connected Car Services" (Paper 4), was presented at the Interdisziplinäre Jahreskonferenz Entrepreneurship, Innovation und Mittelstand (G-Forum) in Karlsruhe (online) on October 1, 2020. The paper "Legal Access to Data as a Dynamic Capability - The Case of Connected Car Services" (Paper 5) was submitted to Strategic Management Journal on October 20, 2021.

The third objective concerns the role of data privacy for digital entrepreneurship and the business opportunities that arise for entrepreneurs from the development of global privacy regulation. While data privacy is often perceived as an obstacle to value-adding processes or as a severe legal risk, the perpetual evolution of data privacy legislation and the need for rule compliance provide business opportunities for start-ups that focus on state-of-the-art digital data privacy solutions. Therefore, the book chapter (*Data are the Fuel for Digital Entrepreneurship - But what about data privacy?*) answers fundamental questions in data privacy areas for digital start-ups and presents different digital entrepreneur opportunities. First, the chapter defines the term data, differentiates it from related areas, and shows what contributes to the value of information. Second, it highlights the significance of data for digital start-ups and demonstrates how start-ups can cope with data protection laws and use the laws to their advantage.

The fourth objective sheds light on the multidisciplinary implications of privacy laws, employing the GDPR as a prime example. It seeks to understand how the research topic of GDPR has progressed within the diverse research domains of law, economics, and computer science, and what insights can be drawn from this progression.

An analysis of the publication volumes of research articles on GDPR in the fields of law, economics, and computer science indicates an overall severe growth of scholarly output in the last decade. As the discourse on GDPR centers around different topics, specific papers and authors are influential in each discipline. There is some indication for a somewhat chronological delay in the academic output on GDPR legislation, where the discourse starts in the field of law and then seems to initiate publications in the fields of computer science and economics.

Considering the impact of present-day data-related legislation in all of the three fields, it is reasonable to ask if this time lag is indicative of lost time in preparing for major changes and if better outcomes can be achieved by engaging in a thorough proactive multidisciplinary examination of data legislation.

Structure of the Thesis

This cumulative thesis consists of two parts: part A establishes a framework for the research papers presented in part B. Part A shows how the thesis' topics build upon fundamental personal rights and freedoms, current legislation, and existing research concerning data processing. The subsequent methodology chapter explains the research approach, the methods used in each study, and motivations for applying the methods. A summary and discussion of the research papers' findings, with an outlook on limitations and further research, conclude Part A.

Part B presents the research papers as published with slight adaptations in format to fit a uniform style within this thesis. The papers can be accessed in their original versions online via their respective publishers.

2. Theoretical Background

The content of this thesis refers to various privacy laws around the world. The following paragraphs describe the common legal and ethical principles, regulations of personal data processing address, which are the background of my empirical research.

2.1. Privacy and Regulation

Article 12 of the Universal Declaration of Human Rights states that "no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks." The European Union introduced a comprehensive framework to guarantee respect for the fundamental rights of individuals to the protection of personal data. Similarly, further guidelines like the Japanese Act on the Protection of Personal Information (APPI), the Personal Information Protection and Electronic Documents Act (PIPEDA) in Canada, or the California Consumer Privacy Act (CCPA) have been developed. The regulations aim at increasing transparency and integrity, and strengthening the position of those concerned in the enforcement of their rights, above all the right to privacy regarding the processing of personally identifiable information (PII), often referred to as personal data.

Personal Data

Article 4 of the GDPR clarifies key terms, defining "personal data" as information relating to an identified or identifiable natural person ("data subject"). Personal data are those, which allow conclusions to be drawn about the actual circumstances of a natural person, i.e., a link between factual information and a person. This may include data that at first glance does not appear to be personally identifiable, for example, a vehicle identification number (VIN). The 'processing' of data is defined as any operation performed on personal data, including the very first steps such as data collection and recording.

Legality and Legitimacy

Article 5 of the GDPR states principles and provides guidelines for processing personal data in two paragraphs. The first paragraph states that personal data must be processed lawfully, fairly, and transparently, and they may only be collected for specified, explicit, and legitimate purposes. In addition, the processing of data must be subject to a limited purpose, which also applies to further processing. According to this principle, data processing must be limited to the necessary extent concerning the purposes for which they are processed. The principle of accuracy requires that personal data must be accurate and kept up to date. Furthermore, it should

only be stored for the period it is required. Finally, the principles of integrity and confidentiality concerning data processing also apply, including appropriate technical and organizational measures to protect the processed data against loss, unlawful processing, destruction, or damage.

The second paragraph describes controllers' obligations to account for compliance with the conditions set out in the first paragraph and to be able to demonstrate that. Personal data must only be processed if legal grounds exist. Furthermore, the processing must be fair and transparent towards the individuals affected. The GDPR explicitly includes the accountability principle, which forces any controller to document and demonstrate compliance with the regulation.

Article 6 of the GDPR summarizes the requirements for the lawfulness of the processing. Paragraph 1 determines conditions, of which at least one has to be fulfilled to justify this. The first stated condition is (a) that the data subject consents to the processing for specific purposes. The lawfulness of the processing is also justified if (b) the performance of contracts, to which the data subject is a party, requires data processing. The same applies to pre-contractual negotiations, which are started at the request of the data subject. Further listed conditions and purposes are, (c) to fulfill a legal obligation of the data subject, (d) to protect the vital interests of the data subject or a third person, (e) to process data in the public interest or the exercise of official authority and (f) to process in the exercise of legitimate interests of the data subject unless fundamental freedoms of the data subject prevail.

Data processing based on legitimate interest is linked to necessary prerequisites. First, the processing must be necessary. The processing would not be lawful if an alternative approach could achieve the same objective. An economic interest is, in itself, not necessarily sufficient. Furthermore, adequate safeguards might be necessary, including pseudonymization and encryption. According to recital 47, such legitimate interests must be weighed against "the interests or fundamental rights and freedoms of the data subject." The controller must prove whether this weighing has been performed and regularly reviewed and must remember that the principles of personal data protection always outweigh the economic interests. The legal grounds must fulfill certain conditions, such as clearly defining the purpose of the data processing.

2.2. Privacy Calculus

The privacy calculus perspective is a widely accepted model for systematically examining willingness to disclose private information (Dinev and Hart, 2006; Krasnova, Spiekermann,

Koroleva, and Hildebrand, 2010; Xu, Teo, Tan, and Agarwal, 2009). Accordingly, consumers perform a risk-benefit analysis of situational factors to decide whether to disclose personal information (Culnan and Armstrong, 1999). If the perceived benefits outweigh the perceived risks, users are willing to share personal information. However, if the risks outweigh the benefits, individuals are restrictive about disclosing data.

The network effects theory examines how an increase in the network size of a user group can create a virtuous cycle (Church & Gandal, 1992; Church et al., 2008; Katz & Shapiro, 1992; Rochet & Tirole, 2003, 2006; Schilling, 2002). In simple terms, it is about situations in which a product or service becomes more valuable when more people use it. Network effects significantly impact users' perceived value of products, services, or platforms (Gregory et al., 2020). Products exhibit network effects when they become more valuable to each user the more people use them (Church & Gandal, 1992; Farrell & Saloner, 1985, 1986; Katz & Shapiro, 1985, 1986, 1992; Sheremata, 2004; Suarez, 2005). Map services with up-to-date traffic information are examples of network effects. The information becomes more accurate and thus more valuable to each user as more people use it and provide traffic data. Research is primarily concerned with two categories: direct and indirect network effects. Direct network effects occur when there is a direct interaction of users on a platform (Rochet & Tirole, 2003; Zhu & Iansiti, 2012). Indirect networks occur when the likelihood of greater availability and variety of additions to the product increases as the number of users increases (Boudreau, 2012; Church, Gandal, Krause, & Canada, 2008; Clements & Ohashi, 2005).

3. Research Methodology

The included papers employ a variety of research methodologies to achieve the research objectives. These approaches extend beyond practical problem analysis and explicitly incorporate systematic literature reviews, bibliometric techniques, and quantitative methods, with a primary focus on multinomial regression analysis.

With the aim of an initial informative practice-oriented, and scientific exploration of the research field, paper 1 is based on the description of a case where data usage poses a practical problem, and paper 2, following on from this, on the conception and description of a possible solution approach operationalizing legal requirements in organizations. While the emphasis in the other papers is on scientific elaboration, paper 1 deliberately explains only the most important fundamentals theoretically and focuses on practical recommendations for action. Based on practical case experience in implementing the requirements of data protection laws at leading German car manufacturers, using the example of transparency and documentation obligations from Article 30 of the GDPR, a logical description was used to identify and describe discrepancies between legislation/regulation and practice. These subsequently indicated research needs for our research and the research community - especially in the practical, regulatory environment. Submissions related to the practice were explicitly solicited at various research conferences to promote exchange between research and practice. In this regard, paper 1 includes a description of the initial situation, objectives, results, and implications for practice, according to the requirements for practice-related submissions.

Paper 2 addresses the challenges described in paper 1 and illustrates a possible optimization of the documentation of processing activities. The focus of the work is on:

- the better implementation of legal requirements within an organization and thus on ensuring compliance
- an increase in efficiency reducing the effort or the use of resources in achieving and ensuring compliance
- the creation of technical possibilities for verifiability, control, or revision of the implementation of requirements.

To this end, based on a review of current research, the development of the Cross-Industry Standard Process for Data Mining model (CRISP-DM) to support compliance through data mining technology was described.

In the third paper, other data-related laws were considered and analyzed in terms of their regulatory purpose. Following this, they were categorized into three categories to bring them into a holistic context. The categorization subsequently enabled the perspectives to be used analytically in practice and thus to achieve a far-reaching view of data compliance within organizations.

The fourth paper formulates practical strategies for integrating personal data into digital product development while ensuring legal compliance and fostering trust. Rooted in a comprehensive analysis of pertinent literature and legal publications, the outlined methodology enables a deep understanding of the value and advantages of consent. These benefits hold particular significance for data-driven businesses as they aid in optimizing targeted data use and mitigating compliance risks. The study includes the formulation of decision-making guidelines, drawn from relevant legal frameworks. This approach encourages the lawful development and enhancement of digital products based on personal data. A distinctive element is the investigation into the differences between consent and legitimate interest as legal bases, as seen through the lens of data privacy practitioners. Data for this inquiry were obtained from an expert survey, which involved both lawyers and certified information privacy professionals. Therefore, we executed a quantitative survey to examine the "flexibility" and "permanence" of consent and legitimate interest, providing valuable insights to aid in the differentiation and evaluation of legal bases in the context of data-driven businesses.

In papers 5 and 6, we contributed to research in the area of strategic management and privacy protection. We tested a research model of the relationships that determine users' general attitudes toward data sharing for CCS. A novel aspect of our research was the explicit consideration of the role of management actions in influencing users' decisions. Using multinomial logistic regression, we found evidence to support our hypotheses (see paper 5). For the analysis, we used a sample of 4,400 individuals from five European countries, for which multivariate statistical techniques had not previously been used. The sample was derived from an EU-wide online survey conducted in August 2017 in which 5,006 individuals (2,430 males and 2,576 females) from Germany, the United Kingdom, France, Italy, and Spain (at least 1,000 individuals per country, 18 years or older) were surveyed. From this original data sample, 560 individuals were eliminated because they did not own a car at the time of the survey and thus could only provide hypothetical answers about their behavior in the relevant context. This resulted in a research sample of 4,440 respondents, whose internal construct consistencies for the independent variables reached very satisfactory values as measured by Cronbach's alpha.

There was no apparent multicollinearity in the sample, as there were no highly correlated independent variables. No outliers were excluded from the analyses, as their exclusion had no evident effect on the results.

The book chapter is based both on relevant literature and on a case practical experience and findings from the management of data risk projects of a leading global accounting firm. No other specific methods were applied to provide answers to fundamental questions in the field of data protection for digital entrepreneurs and to present various options for digital entrepreneurship.

In paper 7 (working paper) we performed a bibliometric analysis of GDPR development in the research disciplines of law, economics, and computer science to reveal the discipline-specific discourses and the most influential works. We show indications of different discipline-specific foci in research interests over time. For practitioners to make well-informed data privacy decisions a multidisciplinary analysis is necessary.

Each method employed in this research offers distinct advantages that contribute to an in-depth understanding of the implications for various stakeholders. The practice-based work provides tangible insights into current challenges and research gaps, drawing from the firsthand experience of requirement implementation in the industry. This approach also proposes solutions to tackle these issues. On the other hand, the systematic literature review assembles and organizes written knowledge to serve as a fundamental reference framework. Meanwhile, empirical research, executed via multinomial regression, sheds light on significant statistical correlations pertinent to the issues at hand. The book chapter presents the researched topics in an organized manner, facilitating knowledge transfer for educational objectives. For practitioners to make well-informed data privacy decisions a multidisciplinary analysis is necessary. These diverse approaches, while each offering unique insights, synergistically complement one another, thereby providing a holistic and enriched understanding of the research subject.

4. Summary of Findings

This chapter summarizes the results of each paper and allocates implications for research and practice. Generally, the findings are broadly consistent throughout the studies and interconnect well to fit the research objectives.

Table 1: Research, Methodologies, Objectives, Main Findings.

#	Full Citation	Methodology	Objective	Main Findings
1	Köhler, W., Schultz, C., Rasche, C. (2020). Das 100% Problem im Datenschutz. 24. Interdisziplinäre Jahreskonferenz zu Entrepreneurship, Innovation und Mittelstand (G-Forum), Karlsruhe.	Practical paper / Insights from practice	To better understand the challenges of implementing privacy legislation	- 100% compliance with GDPR currently difficult to achieve -Processing technologies are a support option
2	Gümüs, C., Köhler, W., Schultz, C.; Rasche, C. (2021). Konzept eines CRISP-DM-Modells zur ganzheitlichen Datenschutzbetrachtung unter Anwendung von Data Mining. In: Reussner, R. H., Koziolek, A. & Heinrich, R. (Hrsg.), INFORMATIK 2021. Gesellschaft für Informatik, Bonn. (1021-1035). ISBN 978-3-88579-708-1.	Literature review and Insights from practice	To better understand possibilities to address challenges of privacy implementation	-100% compliance with GDPR currently difficult to achieve -Data mining as a supporting technology for achieving and maintaining data compliance
3	Köhler, W., Schultz, C., Rasche, C., Herzig, A. (2023, forthcoming). Data Governance Insights – The Magic Triangle of Data Governance.	Practical paper / Insights from practice	To better understand and address the challenges of international data legislation	-Consideration of further data regulations in assessment needed to achieve data compliance and data governance
4	Koehler, W. (2023). Developing Digital Products with Compliance-Driven Personal Data Integration. In: Bitran, I., Bitetti, L., Conn, S., Fishburn, J., Huizing, E., Ritala, P., Torkkeli, M., Yang, J., (Eds.) Proceedings of XXXIV ISPIM Innovation Conference, held in Ljubljana, Slovenia on 04 June to 07 June 2023. ISBN 978-952-65069-3-7.	Literature review, data collection and analysis/ Quantitative research	To better understand and address the challenges of data legislation in digital product development	-Each use case should undergo thorough evaluation for the suitability of legal bases (consent/ legitimate interest) -Decision heuristics effectively guide when and how to use personal data in digital product development.
5	Köhler, W., Schultz, C., Rasche, C. (2020). When Handing out Presents is not Enough! – Influencing Factors on the User´s Willingness to Share Data for Connected Car Services. 24. Interdisziplinäre Jahreskonferenz zu Entrepreneurship, Innovation und Mittelstand (G-Forum), Karlsruhe.	Data collection and analysis/ Quantitative research	To better understand the factors influencing customers' willingness to share personal data and appropriate management measures	-Trust, customer knowledge and transparency influence customer decisions to share data - Appropriate management measures can influence customer decisions
6	Köhler, W., Schultz, C., Rasche, C. (2023, forthcoming). What Determines the Willingness to Share Personal Data?			

	- The Case of the Automotive Industry. *(Submitted to Industry and Innovation)*			
7	Koehler, W., Schultz, C., & Rasche, C. (2022). Data are the fuel for digital entrepreneurship - but what about data privacy?. In: Keyhani, M.,Kollmann, T., Ashjari, A., Sorgner, A., Hull, C. (eds.) Handbook of Digital Entrepreneurship (pp. 306-322). Edward Elgar Publishing. https://doi.org/10.4337/9781800.	Literature review and Insights from practice	To better understand the role of data privacy for digital entrepreneurship	- Development of legislation offers opportunities for digital entrepreneurs, e.g., technical innovations, to meet dynamic processing challenges
8	Köhler, W. (2023, forthcoming, working paper). Mapping the Field Across Disciplines: Data Protection Research in Law, Economics and IT.	Bibliometric analysis	To better understand multidisciplinary implications of privacy legislation	- the fields of law, economics, and IT experienced growth in research interest on GDPR, albeit with a chronological delay - multidisciplinary management of data-related legislation requirements increase efficiency
**The papers are sorted chronologically within the respective research objectives*				

The delineation of the challenges in practice in paper 1 (*Das 100% Problem im Datenschutz*), which mainly deals with implementing the transparency and documentation requirements of the GDPR, shows a great need for technical support for the implementation and maintenance of data compliance. To create and permanently guarantee the necessary transparency, despite the increasing number of processing activities, a growing volume of processed data, and the global development of privacy requirements, technologies for effective and efficient data protection management are receiving increasing attention. In addition, the use of appropriate tools can support risk control and monitoring, compliance auditing, reduce manual efforts and achieve a higher level of compliance.

Paper 2 (*Konzept eines CRISP-DM-Modells zur ganzheitlichen Datenschutzbetrachtung unter Anwendung von Data Mining*) addresses the need for technical support and describes a possible approach to solving the challenges related, using data mining. The model's conception shows that data mining can support the complete collection and categorization of personal data, the derivation of processing activities, and their current documentation. In principle, data mining allows for a holistic, transparent, and centralized control of the "Record of Processing Activities" (RoPA) process, making technological deployment suitable for maintaining compliance. Since the first three phases of the CRISP-DM model take up about 50% – 70% of the workload for developing data mining (Wuttke, 2020), these have ideally already been gone

through by organizations subject to the GDPR. The implementation can build on the current work status and thus shorten it considerably. One recommendation for action concerns the extension of the CRISP-DM model to include a monitoring phase. Data analyses are checked and monitored, e.g., by legal staff, and the analysis is adjusted or corrected as necessary.

In the future, additional sub-processes or further data privacy requirements can be similarly considered and intelligently managed. Research needs may include process mining in combination with data mining. Process mining combines the advantages of data mining with process modeling to make efficient monitoring and the creation of complex real-time processes possible (Reinkemeyer, 2020). By standardizing processes, transparency can be increased, and thus, weaknesses in the current process implementation can be checked and improved (Peters and Nauroth, 2019). Process automation reduces redundancies, avoids bottlenecks, and thus reduces costs (Peters and Nauroth, 2019). An extension of the system to include machine learning may allow it to derive the recommended actions and measures from large data sets of a process (Reinkemeyer, 2020). If the system detects relevant causalities, trends, and patterns based on process-related data and can predict the next activity, it is called predictive process mining.

Paper 3 (*The Magic Triangle of Data Governance*) broadens the view from privacy compliance to other areas of data regulation and data governance. With a focus on practical implications, a model is described that makes it possible to identify and consider additional regulatory perspectives. For this purpose, the international data regulations are re-categorized and their relevance to achieve and validate data governance described. The magic triangles consider all three perspectives of data-related regulations and support the comparison of requirements per jurisdiction and the protection of subjects. This allows existing challenges and inconsistencies in individual laws to be identified, evaluated, and management decisions to be derived from them. Finally, the paper provides management with concrete questions that support the identification of inconsistencies and facilitate the understanding and application of the magic triangle.

Paper 4 (Developing Digital Products with Compliance-Driven Personal Data Integration) emphasizes the crucial role of personal data in digital product development and the importance of responsible data collection and usage. It explores the challenges of transparency and purpose limitation when leveraging vast data sources and big data technologies. The research offers guidance to businesses aiming to balance innovation and privacy in the digital

era, with a detailed analysis of the legal foundations of legitimate interest and consent, and the key factors to consider during their implementation or adaptation to new situations.

For enhancing transparency and user-friendliness in consent processes, the study suggests simplifying language, offering granular consent options, providing timely notices, and ensuring freely given consent. This approach benefits users, service providers, and prompts further research to reconcile user expectations and experiences. To overcome challenges in personal data processing based on legitimate interests, organizations are encouraged to provide balancing test guidelines, improve transparency, strengthen accountability, incorporate privacy by design and default, and uphold data subjects' rights. These strategies ensure regulatory compliance and balance privacy concerns with data processing in digital product development. When changing the legal basis for data processing, organizations are urged to carefully comply with privacy laws and data subjects' rights. The study introduces a decision tree for evaluating data usage in digital product development, ensuring legal compliance, and gathers insights from privacy professionals on GDPR legal bases through an expert survey.

Survey results reveal that consent-based data processing is perceived as more flexible and aligns better with privacy regulation objectives than legitimate interest. Consent allows individuals more control over their data, while legitimate interest necessitates a balancing process by the controller, which is subject to audits or objections. The average preference scores by data management professionals for consent and legitimate interest in aligning with privacy objectives are 8.79 and 6.67, respectively. The results show that a gap between legal bases and their assessment for data-driven businesses needs to be bridged. Overall, this study contributes to a more nuanced understanding of this complex landscape.

Paper 5 (*When Handing out Presents is not Enough! - Influencing Factors on the User´s Willingness to Share Data for Connected Car Services*) deepens the understanding of user willingness to share personal data. The paper initially examines the factors influencing the willingness to share personal data in the context of the use of CCSs. This study demonstrates that different factors determine the user’s willingness. The results of the research model, tested with a multinomial logistic regression, show that the main influential factors regarding the willingness of users to share data for CCSs are:

- knowledgeability about the amount of the data shared
- trust towards the provider
- the perceived personal added value by different areas.

Management measures can potentially influence trust, knowledge about shared data, and perceived personal added value. It becomes clear that simply focusing on added value for the customer, e.g., with gifts such as the free use of services, is not sufficient to ensure the long-term availability of user data. Competitors who want to enter the CCS market might gain a competitive advantage if they educate those critical of sharing their data and seize measures to build trust. Pointing out the advantages can positively influence future decisions of those already willing to share data. Understanding the influencing factors will force new and existing companies to attach greater importance to transparency and communication strategies.

In the context of the 6th paper (*What Determines the Willingness to Share Personal Data? - The Case of the Automotive Industry*), the data from paper 5 were further analyzed, and the factors that influence the willingness to disclose personal data were examined and described more intensively. The results confirm that empirical research efforts around data privacy must be context-dependent. Contextuality is evident in the importance of nearly all data types and data recipients concerning the customers' willingness. Our results also confirm that privacy is a serious issue for users who calculate the benefits and risks of data sharing. Customers expect CCS providers to take privacy seriously. Demographic determinants (e.g., gender, age, country) are of value in that they help channel information about specific target groups, e.g., as users get older, they are less likely to be open to data sharing for CCSs, and men are more likely to be open to data sharing than women. A novel aspect of our research was that we explicitly considered the role of management actions in influencing users' data sharing decisions for CCSs. Understanding the contexts that determine users' data sharing decisions can help dynamically manage users' expectations, perceived needs, and fears. Messages addressing risks and assuring users of a high level of transparency in data collection seem to be valuable approaches in decreasing the risk perceived by users, avoiding consent revocation and motivating users to disclose their data. Continued access to a significant amount of user data is a prerequisite for success in the growing CCSs market. Management needs to alter the benefit/risk calculation toward a perceived positive balance by not only emphasizing the benefits but also thoughtfully addressing perceived risk. The proactive execution of specific management measures can accompany newly designed benefit/risk communication by pointing out that "no data are shared" and "no data triangulation takes place." Our results showed that these management measures were of significant relevance to users. An improved communication strategy and the demonstration of adequate actions to guarantee responsible and lawful data handling might increase trust.

In the book chapter (*Data are the Fuel for Digital Entrepreneurship - But what about data privacy?*), opportunities for digital entrepreneurs to benefit from data privacy regulation are presented. The focus of the book chapter is the importance and advantages of technical compliance innovations. Many opportunities for digital entrepreneurs exist, especially in the area of developing technological solutions for businesses to meet the growing and dynamic challenges of data processing, structuring, valuation, and monetization. Without this support, the underlying conditions creating a permanent need for new products and services may suffocate entrepreneurial creativity.

In our ongoing research paper titled, *" Mapping the Field Across Disciplines: Data Protection Research in Law, Economics and IT"* we highlight the discipline-specific and temporally distinct shifts in research interests concerning GDPR. The disciplines of law, economics, and IT (computer science) have seen a substantial upsurge in research output, although at staggered chronological points. Our analysis of publication volumes underscores a remarkable increase in scholarly discourse in these fields over the last decade. A distinct chronological delay was observed, where discussions about GDPR legislation first manifested in law, followed by IT and economics. Given the profound influence of current data-related legislation in these disciplines, this time lag raises questions about the readiness for significant shifts and indicates the potential advantages of a proactive, multidisciplinary approach concerning data legislation.

5. Conclusion

In the following section, a synthesis of the research findings will be discussed, including limitations of the thesis, open questions, and further research ideas.

Addressing Data Compliance and its Challenges

The first research objective was to better understand operational challenges in implementing data privacy laws and to subsequently investigate suitable technical solutions to improve the implementation of requirements.

It can be concluded that due to the evolving requirements of privacy laws and the rapidly increasing scope of data processing, there is a need for technologies that ensure effective data protection management. These relate both to the implementation of legal requirements and the maintenance of compliance and auditing. Big data technologies such as artificial intelligence will play an increasingly important role in this context in the future, with a wide variety of requirements resulting in a wide variety of research needs and approaches.

One possible approach described in paper 2 is the implementation of a data mining system. Currently, it is very resource- and cost-intensive to meet the transparency and documentation requirements and other requirements of the GDPR. One of the main goals of technical developments in the privacy environment is to increase the efficiency of data privacy processes. Data mining can support implementing the requirements described, but manual activities are still required downstream, especially for legal assessments and for monitoring automatically generated results. By complying with the provisions of Article 30 of the GDPR, the liability risk is significantly reduced. The efficiency of existing processes can be increased but the support level of data mining will only be measurable through the system's real development and actual implementation. Furthermore, the model development is based on assumptions and theories. The CRISP-DM phases have been elaborated in detail but without considering the technological level in depth. Since the technical view is underrepresented in the development of the model, challenges in the implementation can arise that were not fully recognized and considered in the conception. The CRISP-DM model corresponds to a standard model, which serves the standardization of various use cases. Therefore, a generic development of the model was applied in the present work context. In this context, no economic efficiency consideration was performed. Whether implementing a data mining system is worthwhile depends on the organization's size, the costs of compliance, and other factors that have not been considered in detail. The focus of this work was on the transparency and documentation requirements, which stem from Article 30 of the GDPR. The requirements for transparency

about the processing operations form an essential basis for many subsequent data privacy and protection processes. In practice, however, other key challenges would justify a more detailed analysis and research into possible solutions, such as retention and deletion, but which have not been addressed in this paper. Furthermore, the consideration of this thesis has been mainly limited to European legislation and the challenges in implementing its relevant requirements.

There is a need for research around process mining and its combination with data mining. Process mining combines the advantages of data mining with process modeling to make efficient monitoring and the creation of complex real-time processes possible (Reinkemeyer, 2020). In the next step, machine learning can be added to the system. This form of artificial intelligence enables the development of recommended actions and the generation of measures based on large data sets (Reinkemeyer, 2020). This approach is referred to as predictive process mining. Furthermore, academic research should examine the potential of combining process and data mining for standardization efforts and ensuring data privacy requirements in greater depth. Future research shall tackle existing limitations to enrich the multidisciplinary understanding of privacy regulations and optimal practices in digital product development. Data controllers should place a high priority on transparency, informing data subjects about how their data is being used, which would in turn reinforce the lawfulness of processing, mitigate compliance risks, and cultivate trust among customers. The ongoing evolution of digital technologies highlights the necessity for efficient cross-disciplinary collaboration. Further research should aim to perfect existing strategies and develop innovative solutions that reconcile legal requirements, technological progress, and user needs. Promoting interdisciplinary cooperation lays the groundwork for a balanced and efficient framework that aids digital product development, while protecting individual privacy rights and enhancing societal trust in new technologies.

Users' Willingness to Share Personal Data

The second research objective was to gain a better understanding of the factors influencing customers' willingness to share personal data and to investigate how the results help to ensure and expand legal access to personal data. A research model was tested that determines relationships to general user attitudes toward data sharing for CCSs. Papers 4 and 5 show that users' willingness to share personal data depends on various factors, such as trust, knowledge of the scope of the data shared, and perceived added value. The role of management actions in influencing users' data sharing decisions was explicitly considered. Specific management measures can affect these factors and ensure the long-term availability of user data. The results

show that protecting consumers' privacy is a serious matter, and companies are expected to take it seriously and comply with privacy laws.

This study was based on data from five European economies. Results could be different in other markets (USA, China) due to different legislation and general attitudes towards privacy and data protection. Other factors not considered in this study may also play an important role in users' willingness to share data. All variables in this study came from the same questionnaire, so bias from shared methods could have been a problem. However, due to the relatively large sample and clear results, there is no evidence that the results were biased.

Since many services will not be available until the future, customer intentions were the best available proxy. Deriving general conclusions is also complicated by contextuality. Since users respond differently to different stimuli depending on the context, it sometimes leads to contradictory research results. Promising research areas in the area of data privacy include identifying user trust, and empirical analysis of management actions that change users' willingness to share data.

Privacy Paradox

Scientific studies on user behavior regarding data privacy observe that although consumers are concerned about their data security, this concern is not always expressed in concrete action. This observation is also known in the literature as the privacy paradox. It describes the discrepancy between consumers' generally positive attitudes toward data privacy and their actual negligent behavior (Aguirre et al., 2015; Norberg et al., 2007). As Smith et al. (2011) point out, the privacy paradox compromises the results because customers' intentions are not necessarily reflected in their actions.

A rational cost-benefit calculation can explain the discrepancy, i.e., users offset the benefit of specific products or services against the risks of data disclosure and ultimately weigh the benefits of data disclosure higher than the potential risks to their privacy. The basis for this is Behavioral Decision Theory (Kahneman, 2003), according to which users base decisions in complex, uncertain, and risky situations on a rational cost-benefit calculus. (Dinev and Hart, 2004) Situational influences or cognitive biases exist that reduce individual concerns in certain situations. The scientific literature describes many kinds of biases that influence rational decision-making. Various cognitive biases can lead to an irrational and predictable cost-benefit calculation (Simon, 1982). In this regard, Lazarov and Hoffmann (2021) give examples like habituation effects and the related decrease in response to a stimulus (Adjerid et al. 2018; Melumad and Meyer 2020), or the illusion of complete control over data disclosure. (Acquisti

et al. 2013; Martin et al. 2016). In addition to these two perspectives concerning cost-benefit calculation, a third can also unbalance the decision process. Here, prevailing prejudices result in risk assessment not taking place or taking place only to a negligible extent (Barth & de Jong, 2017) so that the value of the desired goal outweighs the risk assessment. An example of this is provided by Shklovski et al., 2014, in which a state of resignation (learned helplessness) is caused by repeated invasion of privacy boundaries.

Outlook

This thesis presents findings pertinent to four primary research objectives: Firstly, it enhances the understanding of the critical operational challenges in implementing data privacy laws. Secondly, it provides insights into factors influencing individuals' readiness to share personal data. As for the third objective, the thesis reveals entrepreneurial opportunities stemming from the evolution of privacy legislation. In addressing the fourth objective, the thesis underscores an analogous surge in research interest in law, economics, and computer science fields, albeit with a time delay. This has implications for a multidisciplinary management approach to data-related legislative requirements aimed at boosting efficiency. The research undertaken in this thesis holistically examines the intricate interplay among various stakeholders, governmental data processing regulations, organizations' implementation of requirements, and user behavior within the context of data privacy.

The study on customer behavior nevertheless leaves some questions open, especially regarding the role of data privacy for product, service, and ultimately company success. The focus of the paper challenges common corporate practices in terms of transparency about the scope, purposes, and technologies used, both in processing and in the protection of personal data. Considering that the current practices of large data-driven companies as Facebook do not necessarily have a direct noticeable effect on the number of users or result in a rethink of the actions of many consumers, the question arises whether data privacy is actually perceived as protecting fundamental rights and whether similar practices under different competitive conditions would lead to severe consequences. In addition, even in a uniform regulatory framework such as the EU, there is still no uniform approach by the authorities to enforce legal claims. The question arises as to whether the necessary resources for the actual and timely processing of infringements in a digital world and whether mechanisms to effectively prevent law dumping through a specific country selection for company representations in the future are already sufficiently available.

Part B – Research Articles

1. Das 100% Problem im Datenschutz

Published: 25. G-Forum Jahreskonferenz (2020), Practice Track - Education and Digitalization

Authors: Wolfgang Köhler, Christian Schultz, Christoph Rasche

Hintergrund

Grundlage dieser Arbeit sind die gesetzlichen Anforderungen gemäß Artikel 30, der europäischen Datenschutzgrundverordnung (DSGVO) nachdem (Abs. 1) jeder Verantwortliche und gegebenenfalls sein Vertreter ein Verzeichnis aller Verarbeitungstätigkeiten führen, die ihrer Zuständigkeit unterliegen. Des Weiteren sind nach Artikel 30 DSGVO Abs. 2 Auftragsverarbeiter und gegebenenfalls Vertreter dazu verpflichtet, ein Verzeichnis zu allen Kategorien von im Auftrag eines Verantwortlichen durchgeführten Tätigkeiten der Verarbeitung zu führen.

Das Verzeichnis ist schriftlich zu führen, was auch in einem elektronischen Format erfolgen kann (Abs. 3). Verantwortliche oder Auftragsverarbeiter sowie gegebenenfalls deren Vertreter sind dazu verpflichtet, der Aufsichtsbehörde das Verzeichnis auf Anfrage zur Verfügung zu stellen (Abs 4).

Die gennannten Pflichten gelten für alle Unternehmen, die nicht ausdrücklich durch Abs. 5 ausgenommen sind.

Ausgangssituation

Verschiedene Datenschutzgesetze verlangen, dass Unternehmen detaillierte Informationen über Praktiken der Datenerhebung, Datenverarbeitung, Weitergabe und Speicherung dokumentieren sowie an die betroffenen Personen kommunizieren. Im Falle der europäischen Datenschutz-Grundverordnung (DSGVO) müssen alle Verarbeitungsprozesse, die personenbezogene Daten betreffen, vollständig und aktuell in einem Verzeichnis der Verarbeitungsaktivitäten (VVT) dokumentiert werden.

In jedem Fall müssen die Unternehmen Transparenz hinsichtlich erhobener und verarbeiteter Daten, der Verarbeitungsverfahren und –systeme sowie der internen und externen Datenströme herstellen und aufrechterhalten. Die Betroffenen müssen vor der Verarbeitung (gem. DSGVO) oder auf Anfrage (gem. verschiedener US Privacy Bills) über den Umfang und die Zwecke der Verarbeitung informiert werden. Alle relevanten Verarbeitungsprozesse, inklusive der verarbeiteten Datenarten und Zwecke sind im VVT zu dokumentieren. Das Verzeichnis kann jederzeit von den Datenschutzbehörden angefordert und geprüft werden. Die Nachweispflicht liegt bei den Unternehmen.

Insbesondere in größeren, international agierenden Unternehmen sind die jeweiligen Entitäten in der Pflicht, Transparenz zu schaffen und Dokumentationsanforderungen umzusetzen. In der Praxis werden Dokumentations- und Aktualisierungsanforderungen auf Bereichs- oder Abteilungsebene heruntergebrochen. Es werden Rollen wie Prozess- und Dateneigner definiert, die für die Einhaltung der gesetzlichen Vorgaben hinsichtlich der Verarbeitungsprozesse im jeweiligen Verantwortungsbereich zuständig sind. Die Erfassung und Aktualisierung von Verarbeitungsprozessen werden häufig abteilungsintern und somit dezentral umgesetzt. Unterstützt werden die Fachbereiche von eingesetzten Datenschutzbeauftragten oder Datenschutzkoordinatoren. Die Ergebnisse der dezentralen Erfassung oder Aktualisierung fließen schließlich in ein zentrales Verarbeitungsverzeichnis ein

Problemstellung

Verarbeitungsprozesse sind vielfältig, dynamisch und im ständigen Wandel. Um Aktualität sicherstellen zu können, sind laufende Anpassungen notwendig. Einige Verfahren sind über einen

längeren Zeitraum hinweg beständig, andere ändern sich häufig. Ein regelmäßiger Turnus zur Prüfung der Aktualität ist somit nur bedingt geeignet.

Verarbeitungszwecke und –Techniken sind ebenfalls dynamisch. Insbesondere bei der (Weiter-) Entwicklung digitaler Produkte und dem steigenden Einsatz neuer Technologien in der Datenverarbeitung, steigen die Verarbeitungsmöglichkeiten. Den Betroffenen vorab, vollständig über Umfang und Zwecke zu informieren steht einer stetigen Entwicklung in der Verarbeitung entgegen.

Zudem wird Verarbeitung personenbezogener Daten wird nicht zwangsläufig als solche identifiziert. Dies ist der Fall, wenn Datenkategorien verarbeitet werden, die ohne konkrete Kenntnisse nicht als personenbezogene oder personenbeziehbare Daten erkannt werden (z.B. Fahrzeug-Identifizierungsnummer). Weiterhin besteht eine Heraus-forderung in der Identifikation von personenbezogenen Daten, die sich aus einer Kombination verschiedener, isoliert betrachtet unkritischer Daten, ergeben kann. Darüber hinaus ist die Vollständigkeit (100%) des VVTs zumeist unbekannt. Es ist nicht möglich eine „Soll-Situation“ zu definieren, wenn der betreffenden Organisation, die Gesamtheit der existenten Datenverarbeitungsprozesse nicht bekannt ist. Somit kann weder die vollständige Umsetzung noch der Erfüllungsgrad der Anforderungen umfassend geprüft werden. Insbesondere gilt dies für größere Unternehmen mit verschiedenen Entitäten und Geschäftsbereichen.

Grundlegende Fragestellungen in der Organisation:

- Was sind personenbezogene Daten und in welchen Prozessen werden diese verarbeitet?
- Auf welchen Rechtsgrundlagen der Verarbeitung basieren die Verarbeitungsprozesse?
- Welche Verarbeitungszwecke liegen vor und wann entfallen diese bzw. wann muss gelöscht werden?
- Für welche personenbezogenen Daten besteht eine Aufbewahrungspflicht und wie lange müssen die betreffenden Daten gespeichert werden?
- Kennt jeder Mitarbeiter die rechtlichen Vorgaben und die operativen Auswirkungen sowie Handlungsbedarfe

- Ist jeder Mitarbeiter in der Organisation in der Lage, die relevanten Prozesse zu identifizieren, dokumentieren und Handlungsbedarfe abzuleiten? (inkl. Betriebsarzt, Empfang, Personalwesen, etc.)
- Sind Verantwortlichkeiten (ins. DPO, CISO etc.) und der Verantwortungsübergang klar geregelt?
- Kann die Vorlagefähigkeit unter Einhaltung der Anforderungen gemäß Art. 30 DSGVO, gewährleistet und aufrechten werden?

Auswirkungen in der Praxis

Das Verzeichnis der Verarbeitungstätigkeiten wird häufig mit großem manuellem Aufwand erstellt. Dezentral erhobene und beschriebene Verfahren werden in einer zentralen Datei dokumentiert. Es gibt Bestrebungen, dieses Vorgehen zentral zu steuern, zu unterstützen und nachzuverfolgen. Es bestehen offizielle sowie häufig unternehmensinterne Leitlinien und beschriebene Dokumentationsanforderungen. Teilweise kommen unterstützend auch Tools und Systeme zum Einsatz (z.B. Privacy Management Tools und Software). Die Nachverfolgung und Prüfung erfolgt nachgelagert, ohne dabei den Soll-Zustand hinsichtlich Vollständigkeit und Aktualität zu kennen.

Prüfungshandlungen können Abweichungen identifizieren, allerdings:

- werden nur Abweichungen identifiziert, die im Prüfungsprogramm explizit enthalten sind. Ohne Kenntnis über den Soll-Zustand (die 100%), ist ein Soll/Ist Abgleich nicht möglich.
- spiegeln die Ergebnisse lediglich den Stand des Prüfungszeitpunktes wider. Durch Prüfungen im üblichen Audit-Rhythmus können Risiken und Handlungsbedarfe weder zeitnah noch lückenlos identifiziert werden.
- ist das Audit nachgelagert und ersetzt nicht die von der Unternehmensleitung eingerichteten Risikokontrollen zur Einhaltung gesetzlicher Vorschriften.
- ist eine manuelle Prüfung von verschiedenartigen Prozessen mit verschiedenen Merkmalen von bis zu mehreren tausend Verarbeitungsprozessen sehr aufwändig.
- erfordert jede neue Entwicklung in der Verarbeitung eine Prüfung und ggf. erneute Umsetzung von Informationspflichten und Anpassung der Prozessdokumentation.

- stehen heterogene Prozesse und vielfältige Systemlandschaften einem zentralen, einheitlichen Prüfprogramm entgegen.
- können Kompetenzen, Kapazitäten und Prioritäten innerhalb einer Organisation stark voneinander abweichen.

Implikationen für Forschung und Praxis

Um trotz der steigenden Anzahl an Verarbeitungsprozessen und des wachsenden Umfangs verarbeiteter Daten, Transparenz herzustellen und dauerhaft zu gewährleisten zu können, wächst der Bedarf an neuen Technologien für ein effektives Datenschutzmanagement. Es bedarf Managementkontrollen sowie geeignete Instrumente zur Risikokontrolle und Überwachung hinsichtlich der Einhaltung von Vorschriften (1st & 2nd line of defense). Systemische Unterstützung bei der Erhebung, Dokumentation, Aktualisierung sowie Prüfung erforderlich, um die enormen manuellen Aufwände zu reduzieren und einen höheren Erfüllungsgrad erreichen zu können

Dabei rücken neue Tools und Technologien wie Data -und Process Mining, die es ermöglichen Verarbeitungsprozesse und Datenströme zu visualisieren, weiter in den Fokus. Der Einsatz von künstlicher Intelligenz, Algorithmen und Systemen zur Abbildung und laufenden Kontrolle von Daten im Unternehmen kann sowohl die Konformität als auch die Effizienz erheblich unterstützen. Abweichungen und Veränderungen innerhalb von Datenverarbeitungsprozessen können erkannte und Risiken sowie Handlungsbedarfe effizient abgeleitet werden. Künstliche Intelligenz kann zur Erkennung von Verarbeitungsprozessen auf Basis verschiedener Merkmale eingesetzt werden. In weiteren Ausbaustufen können Möglichkeiten zur laufenden Identifikation von Veränderungen in den Verarbeitungsprozessen, zur Prüfung der Umsetzung fachlicher Löschkonzepte anhand verschiedener Merkmale, zur Identifikation von Verarbeitungszwecken, den Abgleich mit den umgesetzten Informationspflichten sowie Rechtsgrundlage der Verarbeitung, entwickelt werden.

Ein globaler, industrieübergreifender Einsatz ist möglich. Die Durchführung von Prüfungshandlungen als auch die unternehmensinterne Umsetzung in Form eines Daten-schutzmanagement-Systems, zur Reduktion und Kontrolle von Compliance-Risiken im Unternehmen wird angestrebt. Um der beschriebenen Problemstellung angemessen zu begegnen, bestehen Forschungsbedarfe und vielfältige Forschungsansätze, insbesondere bei der Entwicklung geeigneter technischer Lösungen.

2. Konzept eines CRISP-DM-Modells zur ganzheitlichen Datenschutzbetrachtung unter Anwendung von Data Mining

Abstract

Während die Digitalisierung weiter voranschreitet und immer größere Datenmengen verarbeitet werden, müssen zeitgleich steigende gesetzliche Anforderungen im Umgang mit Daten, insbesondere zum Schutz der Rechte und Freiheiten natürlicher Personen, beachtet werden. Um die gesetzliche Konformität von Datenverarbeitungsprozessen sicherzustellen, sind Organisationen in der Pflicht, Transparenz über Verfahren zur Erfassung und Verarbeitung personenbezogener Daten herzustellen. Unternehmen greifen zunehmend auf innovative Datenanalytik-Technologien zurück, um Analysen großer Datenmengen durchführen zu können und Muster von oder Verbindungen zwischen Daten zu erkennen. Der Beitrag nimmt sich der Optimierung des Dokumentations- und Aktualisierungsprozesses von Verarbeitungstätigkeiten an und be-fasst sich mit der Entwicklung des Cross-Industry Standard Process for Data Mining Modells (CRISP-DM) zur Wahrung der Konformität durch den Einsatz von Data Mining. Einleitend wird der Stand der Wissenschaft und die Methodik zur Modellentwicklung dargelegt, worauf-hin die einzelnen Phasen des CRISP-DM konzipiert werden.

Keywords: Datenschutz; Datenschutzgrundverordnung; Personenbezogene Daten; Verarbeitungstätigkeiten; Digitalisierung; Data Mining; Künstliche Intelligenz; Cross-Industry Standard Process for Data Mining, CRISP-DM

Published: 51st Annual Conference of the German Informatics Society (INFORMATIK 2021), Panel Recht und Technik (RuT2021), Berlin

Authors: Can Gümüs, Wolfgang Köhler, Christian Schultz, Christoph Rasche

Einführung

Der fortschreitende Wandel hin zu digitalen Geschäftsmodellen und Arbeitsprozessen macht die Erhebung stetig wachsender Datenmengen notwendig. Gleichzeitig treten immer mehr Gesetze zum Schutz der Privatsphäre natürlicher Personen und deren personenbezogenen Daten in Kraft. Organisationen sehen sich zunehmend damit konfrontiert, detaillierte Informationen über Praktiken der Datenerfassung und -verarbeitung zu dokumentieren und Transparenz hinsichtlich interner Verfahren im Umgang mit Verarbeitungsprozessen sicherzustellen. So sind etwa Umfang und Zweck einer Verarbeitung vor Beginn der Datenverarbeitung gegenüber betroffenen Personen offenzulegen (Art. 13 DSGVO). Verarbeitungsprozesse sind in einem zentralen Verzeichnis der Verarbeitungstätigkeiten (VVT) festzuhalten (Art. 30 Abs. 1 DSGVO). Das Verzeichnis kann zu jeder Zeit von Datenschutzbehörden angefordert werden, wobei die Nachweispflicht dem Unternehmen obliegt (Art. 30 Abs. 4 DSGVO). Ausgenommen von den genannten Pflichten sind Organisationen mit weniger als 250 Beschäftigten – unter der Voraussetzung, dass durch die Verarbeitung kein Risiko für die Rechte und Freiheiten der Betroffenen besteht, die Verarbeitung nur gelegentlich erfolgt oder keine besonderen Datenkategorien gemäß Artikel 9 Absatz 1 oder Artikel 10 DSGVO verarbeitet werden (Art. 30 Abs. 5 DSGVO).

Die Praxis zeigt, dass insbesondere große, international operierende Organisationen die Umsetzung von Dokumentations- und Aktualisierungsanforderungen zu Verarbeitungstätigkeiten häufig auf Bereichs- oder Abteilungsebene herunterbrechen (Köhler et al., 2020). Eine grundlegende Herausforderung betrifft die Sicherstellung von Vollständigkeit und Aktualität des zentralen VVT. Mangels fehlender Transparenz ist für viele Unternehmen die Gesamtheit der existierenden Verarbeitungsprozesse unbekannt. Somit sind weder die Ganzheitlichkeit noch der Erfüllungsgrad gesetzlicher Anforderungen in vollem Umfang überprüfbar. (Köhler et al., 2020). Es fehlt eine Übersicht aller verarbeiteten personenbezogenen Daten, damit Verantwortliche der Datenverarbeitung die rechtskonforme Umsetzung datenschutzrechtlicher Vorgaben prüfen können. Überdies existieren meist keine einheitlichen Standards zur Dokumentation von Verarbeitungstätigkeiten. Daher sind Verarbeitungsprozesse oftmals heterogen organisiert und unvollständig oder fehlerhaft dokumentiert. Eine weitere Problemstellung resultiert aus der Dynamik und Vielfalt von Verarbeitungsprozessen. Während einige Verfahren beständig sind, befinden sich andere in einem ständigen Wandel. Ein regelmäßiger Turnus zur Aktualitäts- und Konformitätsprüfung ist daher nur bedingt geeignet (Köhler et al., 2020). Sofern Datenkategorien

verarbeitet werden, die ohne konkrete Kenntnis nicht als personenbezogene Daten identifizierbar sind, kann dies maßgeblich die Sicherstellung der Datenschutzkonformität beeinflussen (Köhler et al., 2020). Werden personenbezogene Daten nicht als solche identifiziert, erfolgt auch keine Überführung und Zentralisierung der betroffenen Verarbeitungstätigkeiten im VVT. Im Falle einer behördlichen Prüfung drohen Unternehmen hohe Geldstrafen sowie Reputationsschäden.

Um den Problemstellungen entgegenzuwirken und Konformität zu gewährleisten, greifen Organisationen vermehrt auf Technologien wie etwa Data Mining zurück. Sie ermöglichen die nahtlose Analyse großer Datenmengen, transparente Visualisierungen und Überwachung von Datenströmen und lassen zusammenhängende Muster und Abhängigkeiten zwischen Daten erkennen (Hackett, 2016). Im Fokus des vorliegenden Beitrags steht die Frage, wie Data Mining zur Einhaltung des Datenschutzes in Unternehmen entwickelt werden kann.

Ein besonderes Augenmerk liegt auf der Pflege des VVT, da dieses das zentrale Element der europäischen Datenschutzgrundverordnung (DSGVO) darstellt. Es wird geprüft, welche technologischen Anpassungen zur Lösung der Problemstellungen essentiell sind und wie der Einsatz von Data Mining effizient umgesetzt, die Komplexität der Arbeitsvorgänge verringert und die Flexibilität von Geschäftsprozessen erhöht werden kann.

Verwandte Arbeiten

Der folgende Abschnitt befasst sich mit einer Vorstellung bereits existierender Vorgehensmodelle zur Sicherstellung datenschutzrechtlicher Vorgaben, die auf dem Einsatz innovativer Technologien der Datenanalytik beruhen.

Im Beitrag von Becker und Buchkremer wird die Entwicklung eines agilen Vorgehensmodells erläutert, mit dessen Hilfe aufsichtsrechtliche Anforderungen durch Einsatz einer sogenannten Regulatory Technology Lösung implementierbar seien (Becker & Buchkremer, 2018). Die Autoren heben die Relevanz eines harmonischen Zusammenspiels zwischen Technologie und menschlichen Experten für agile Implementierungsprozesse hervor und kommen zu dem Schluss, dass iterative Vorgehen für die Analyse regulatorischer Anforderungen im Kontext des Datenschutzes erfolgsentscheidend sind.

Kittel beschreibt in einem Artikel, wie Agilität bei Geschäftsprozessen mit Datenschutzbezug sichergestellt werden kann (Kittel, 2013). Es zeigt sich, dass Ad-hoc-

Änderungen von Geschäftsprozessen dieser Art eine vorausgehende Kontrolle regulatorischer Datenschutzanforderungen unbrauchbar machen. Kittel stellt einen modellbasierten Ansatz zur Ad-hoc-Integration von Datenschutzkontrollen in Arbeitsabläufen vor, durch den die Abhängigkeiten zwischen Agilität und Compliance verringert werden sollen.

Ein weiteres Vorgehensmodell zur Vorbereitung auf datenschutzrechtliche Anforderungen wird von Wirnsperger, Buchholz und Wolff erarbeitet (Buchholz et al., 2016). Das Modell berücksichtigt rechtliche, technische, organisatorische und prozessuale Aspekte. Beginnend mit einer Vorprüfung und einer Umfeldanalyse zur Erfassung aller personenbezogenen Daten in Geschäftsprozessen solle der Status Quo auf Basis einer Fit-/Gap-Analyse erfasst sowie ein Risiko- und Maßnahmenplan erarbeitet werden.

Das von Chapman et. al. entwickelte CRISP-DM-Modell stellt die Entwicklung und Umsetzung spezifischer Data-Science-Projekte durch den Einsatz von Data Mining und künstlicher Intelligenz in den Mittelpunkt (Chapman et al., 2000). Das Modell gilt als Standardvorgehen für die Ausführung von Data-Mining-Projekten und ist für diverse Projekte der künstlichen Intelligenz zur Sicherstellung des Datenschutzes anwendbar. Da das CRISP-DM-Modell in den Kontext des aktuellen Technologiestands eingeordnet ist, wird es als Rahmen für die vorliegende Untersuchung verwendet.

Während ein Großteil aktueller Untersuchungen den Einfluss regulatorischer Datenschutzvorgaben auf die Entwicklung intelligenter Technologien diskutieren, widmen sich einige wenige Quellen dem Unterstützungsgrad innovativer Technologien und deren Anwendungspotentialen zur Wahrung des Datenschutzes. Inwiefern Data Mining jedoch speziell bei der Verarbeitungsdokumentation und -aktualisierung in einem VVT unterstützt, wird in der Wissenschaft nicht vertieft betrachtet. Nach aktuellem Stand existiert kein Vorgehensmodell für diesen spezifischen Anwendungsfall.

Vorgehensmodell und methodische Unterstützung

Anforderungsanalyse

Die Anwendung einer Anforderungsanalyse hat unmittelbaren Einfluss auf die zielgerichtete Entwicklung des CRISP-DM-Modells. Zur vollständigen Ermittlung aller Anforderungen an CRISP-DM wird zunächst ein umfassender Anforderungskatalog entwickelt. Der Katalog

differenziert zwischen technologischen Anforderungen, die primär die zu erbringenden Funktionalitäten, Mechanismen und Leistungen des Data Mining zur Gewährleistung der Konformität betreffen und Anforderungen von Seiten des Datenschutzrechts zur Pflege eines zentralen VVT. Letzteres orientiert sich an den Regularien der DSGVO aus Artikel 30.

CRISP-DM

CRISP-DM folgt einem iterativen Kreislauf mit insgesamt sechs Phasen, ohne dabei einen konkreten Endpunkt festzulegen. Stattdessen kann jede Phase und deren Iterationen, je nach Problemstellung mehrfach durchlaufen und ausdifferenziert werden. Jede Wiederholung des Gesamtprozesses bringt neue Fragestellungen hervor und kann zu einer Prozessoptimierung beitragen. Das Modell schreibt keine starre Sequenzierung der einzelnen Phasen vor. Rückkopplungen, die sich etwa aus unvorhergesehen Problemfaktoren oder unzureichender Qualität eines Zwischenergebnisses ergeben, sind durchaus möglich und gewünscht. (Chapman et al., 2000)

Phase 1. Zur Erlangung vollständiger Kenntnis über die Geschäftsanforderungen und konkrete Aufgabenstellung hat eine präzise Erörterung der betriebswirtschaftlichen Problemstellung zu erfolgen (Chapman et al., 2000). Dabei sind die zu erreichenden Zielkriterien festzulegen. Diese werden in Anforderungen an die Datenanalyse überführt, woraufhin ein konkreter Umsetzungsplan unter Berücksichtigung zeitlicher, personeller und sachlicher Ressourcen aufzusetzen ist (Cleve & Lämmel, 2016).

Phase 2. Im nächsten Schritt werden relevante Datenbestände selektiert, deren Verarbeitung zur Erfüllung der zuvor bestimmten Ziele notwendig ist (Cleve & Lämmel, 2016). Es wird eine Datensammlung mit Beschreibung der typischen Eigenschaften der relevanten Daten angelegt, um ein generelles Verständnis über die selektierten Daten aufzubauen. Die Phase mündet letztlich in einer Bewertung der Datenqualität und -quantität (Cleve & Lämmel, 2016).

Phase 3. Die Datenvorbereitung zielt auf die Auswahl der finalen Datenmenge ab, die in das Data-Mining-System integriert und entlang vordefinierter, anwendungsspezifischer Algorithmen analysiert werden soll (Chapman et al., 2000). Es bedarf einer klaren Differenzierung zwischen irrelevanten und relevanten Daten. Das Ergebnis der Datenauswahl hängt von der jeweiligen Zielsetzung des Data-Science-Projektes ab. Ferner sind die Daten zu bereinigen, um eine Data-Mining-Verarbeitung zu ermöglichen. Diese Phase entscheidet darüber, welche speziellen

Merkmale und Charakteristiken die nachfolgende Modellbildung berücksichtigen soll (Cleve & Lämmel, 2016).

Phase 4. Die Modellbildung nimmt sich der eigentlichen Datenanalyse an, indem ein Modell zum Umgang mit den selektierten Daten entwickelt wird (Cleve & Lämmel, 2016). Nach Auswahl und Parametrisierung einer passenden Modellierungstechnik wird ein Testmodell entwickelt, mit dessen Hilfe die Präzision und Qualität des Entwicklungsergebnisses geprüft und bewertet wird. Die Algorithmen der Modellbildung unterscheiden zwischen einem *Trainieren* und *Anwenden*, wobei das Modell entweder auf Basis des gewonnenen Wissens aus historischen Daten trainiert oder auf neue, bisher unbekannte Datensätze angewendet wird.

Phase 5. Zur Evaluation des Entwicklungsergebnisses wird die eingangs festgelegte Zielsetzung mit dem erarbeiteten Data-Mining-Verfahren abgeglichen. Für den Fall, dass die gewünschte Qualität des Modells zur Erfüllung der Zielkriteren nicht vollständig oder nur in Teilen erreicht wurde, muss CRISP-DM erneut durchlaufen werden (Cleve & Lämmel, 2016).

Phase 6. Den Abschluss bildet die Planung und Umsetzung der Implementierung des Data Mining im Unternehmen. Das Modell kann je nach Anwendungsfall auf existierende oder auf neue, bislang unbekannte Datenbestände angewendet werden (Chapman et al., 2000).

Konzipierung des CRISP-DM-Modells

Anforderungsspezifizierung

Technologische Anforderungen. Tabelle 1 zeigt einen Überblick der technologischen Anforderungen an das Data-Mining-System.

Table 2: Technologische Anforderungen an das Data-Mining-System.

Technologische Anforderungen
Zugriff auf den gesamten Datenpool des Unternehmens
Erschließen aller im Unternehmen verfügbaren, (un-)strukturierten Daten
Identifikation und Strukturierung aller verarbeiteten personenbezogenen Daten
Erfassung aller existierenden Verarbeitungsprozesse

Technologische Anforderungen
Zentrale Steuerung der Pflege eines VVT
Vollständige Dokumentation aller Verarbeitungsprozesse im VVT
Gewährleistung kontinuierlicher Aktualität des VVT
Automatische Anpassung und Aktualisierung von Verarbeitungsprozessen
Erkennen von Trends, Veränderungen und datenschutzrechtlichen Anforderungen

Datenschutzrechtliche Anforderungen. Tabelle 2 zeigt einen Überblick der datenschutzrechtlichen Anforderungen an die Dokumentation von Verarbeitungstätigkeiten in einem VVT gemäß Artikel 30 Absatz 1 DSGVO. Gleiches gilt für Auftragsverarbeiter unter Ausschluss der Beschreibung und Kategorisierung der Verarbeitungszwecke, der Beschreibung und Kategorisierung aller Datenempfänger im In- und Ausland sowie der Löschfristen der verschiedenen Datenkategorien (Art. 30 Abs. 2 DSGVO).

Table 3: Datenschutzrechtliche Anforderungen.

Datenschutzrechtliche Anforderungen
Name und Kontaktdaten des verantwortlichen Datenverarbeiters
Beschreibung und Kategorisierung der Verarbeitungszwecken
Beschreibung des Betroffenen und Kategorisierung der betroffenen Personen
Beschreibung und Kategorisierung der personenbezogenen Daten des Betroffenen
Beschreibung und Kategorisierung aller Datenempfänger im In- und Ausland
Beschreibung der Übermittlung in Drittländer oder internationale Organisationen und deren Benennung
Löschfristen der verschiedenen Datenkategorien
Dokumentation der technischen und organisatorischen Maßnahmen

Grundlagen

Die erforderlichen Inhalte eines VVT ergeben sich aus der Analyse manuell gepflegter Verzeichnisse in der Praxis. Dies dient der nachgelagerten Lösungssuche, indem ermittelt wird, welche Strukturierungen und Klassifizierungen der relevanten Daten das Data-Mining-System zur vollumfänglichen Dokumentation zu berücksichtigen hat. Im Kontext der CRISP-DM-Entwicklung werden schließlich Regeln, Korrelationen und Muster zwischen Daten und deren Verarbeitungstätigkeiten abgeleitet. Eine beispielhafte Übersicht des Aufbaus eines VVT und der zu dokumentierenden Inhalte und Informationen ist in Tabelle 3 gegeben.

Table 4: Themenbereiche und potentielle Inhalte eines VVT.

Themenbereiche eines VVT	Potentielle Inhalte
Dokumentation der Kontaktdaten des verantwortlichen Daten- oder Auftragsverarbeiters	Name, Funktion, E-Mail-Adresse, Telefonnummer und Anschrift des Verantwortlichen oder Auftragsverarbeiters

Themenbereiche eines VVT	Potentielle Inhalte
Dokumentation der Verarbeitungsprozesse	Bezeichnung, Beschreibung, Datenherkunft, Verwendetes IT-System
Dokumentation des Zwecks der Datenverarbeitung	Zweckkategorie, Zweckänderung, Zweck (Mit-)Bestimmung durch Dritte
Dokumentation des datenverarbeitenden Systems	Name des datenverarbeitenden Systems
weitere	

Konzeption

Phase 1. Zur Generierung eines exakten Verständnisses der Aufgaben- und Problemstellung werden im ersten Schritt die erwarteten Projektziele sowie -ergebnisse festgelegt. Im vorliegenden Kontext leiten sich diese aus den technologischen und datenschutzrechtlichen Anforderungen ab. Tabelle 4 zeigt die potentielle Zielsetzung des Data-Mining-Vorhabens.

Table 5: Zielformulierung.

Zielformulierung
Steigerung der Transparenz
Steigerung der Effizienz bei Anpassungen an Verarbeitungsprozesse
Verfolgbarkeit bei Prozessaktivitäten
Verfügbarkeit relevanter Informationen
Reduzierung der Arbeitsauslastung von Fachbereichen eines Unternehmens
Reduzierung des Abstimmungsaufwands zwischen den Fachbereichen
Standardisierung des Vorgehens
Vollständige Identifikation und Strukturierung personenbezogener Daten
Vollständige Erfassung aller Verarbeitungsprozesse
Unterstützung der Dokumentation der Verarbeitungsprozesse im VVT
Sicherstellung kontinuierlicher Aktualität des VVT

Weiterhin ist eine ausführliche Risikoanalyse durchzuführen (Cleve & Lämmel, 2016). Ermittelte Risiken sind gemäß projektspezifischen Kriterien zu bewerten und individuell zu analysieren. Dies kann mittels einer Risiko-Matrix erfolgen, wobei die Eintrittswahrscheinlichkeit und Schadenhöhe für jedes Risiko geschätzt und in der Risiko-Matrix visualisiert werden. Zur Reduktion besonders schwerwiegender Risiken sind Gegenmaßnahmen festzulegen. Das Risikomanagement hat über die gesamte Projektdauer hinweg zu erfolgen.

Anhand der Zielsetzung werden Erfolgskriterien zur finalen Bewertung des Entwicklungsergebnisses spezifiziert (Chapman et al., 2000). Das Vorhaben ist dann erfolgreich, wenn die Gesamtheit aller personenbezogenen Daten bekannt ist und diese entlang charakteristischer Merkmale strukturiert werden. Entsprechend sind Verarbeitungstätigkeiten

automatisch zu erfassen, sodass Mitarbeitende bei Zentralisierung und Aktualisierung des VVT unterstützt werden. Weitere Erfolgskriterien betreffen die Reduzierung der Fehleranfälligkeit, des manuellen Arbeitsaufwands, der Komplexität sowie der Arbeitsauslastung innerhalb der Fachbereiche und -abteilungen.

Die Festlegung von Unternehmenszielen orientiert sich vor allem an der Firmenkultur und Vision eines Unternehmens, weshalb eine allgemein gültige Aussage nur bedingt möglich ist. Nichtsdestominder spiegeln sich zumeist einige Unternehmensziele wie Integrität, Verlässlichkeit und Vertrauenswürdigkeit in verschiedenen Unternehmen wider und gehen demnach mit den Grundsätzen und Schutzzielen der Datensicherheit und des Datenschutzes (Compliance) einher. Die strikte Einhaltung dieser drei Faktoren sind in der heutigen Zeit zur Wahrung der Wettbewerbsfähigkeit (Hellmann, 2018) und Rechtmäßigkeit unabdingbar.

Die wichtigsten Fragestellungen zur Aufbereitung der ersten Phase sind nachfolgend zusammengefasst.

- Wie sind Ausgangssituation und Problemstellung?
  - Welche Ziele und Ergebnisse sollen durch Data Mining erreicht werden?
  - Welche Risiken (finanziell, rechtlich, organisatorisch) können auftreten?
  - Was sind die Erfolgskriterien und Unternehmensziele?
  - Wie ist die aktuelle Unternehmenssituation?
  - Welche Ressourcen sind zur Umsetzung des Vorhabens verfügbar?
  - Welche Kosten sind für welchen Nutzen aufzubringen?
  - Wurde ein Projektmanagementsystem etabliert und ein Projektplan aufgesetzt?

Phase 2. Zur Selektion der relevanten Datenbestände muss geklärt werden, was unter personenbezogenen Daten verstanden wird und wodurch sich diese kennzeichnen. Die DSGVO definiert personenbezogene Daten als „Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person beziehen“ (Art. 4 Abs. 1 DSGVO). Personen gelten dann als „identifizierbar“, wenn sie sich (in-)direkt eindeutig identifizieren lassen (Art. 4 Abs. 1 DSGVO). Es wird dann von personenbezogenen Daten gesprochen, wenn die erhobenen Daten einen direkten Bezug zu einer betroffenen Person hervorbringen.

Die Elemente und Kategorien der personenbezogenen Daten sind zu ermitteln. Dies erlaubt die Auswahl und Entwicklung eines passenden Data-Mining-Vorgehens. Datenelemente und -

kategorien leiten sich aus den in Tabelle 3 dargelegten Informationen zur Dokumentation von Verarbeitungstätigkeiten ab. Die Erkenntnisse dienen der Entwicklung von Mustern und Regeln, die Data Mining zur vollumfänglichen Identifikation der personenbezogenen Daten, zur Ableitung und Dokumentation der resultierenden Verarbeitungstätigkeiten sowie dem Segmentieren der Elemente und Kategorien anzuwenden hat.

Verantwortliche mit weniger als 250 Beschäftigten sind gemäß Artikel 30 Absatz 5 DSGVO dazu verpflichtet, zu erheben, ob mit der Verarbeitung der personenbezogenen Daten besondere Risiken für die Rechte und Freiheiten für die Betroffenen einhergehen und inwieweit besondere Kategorien personenbezogener Daten gemäß Artikel 9 DSGVO verarbeitet werden. Weiterhin ist der Turnus der Datenverarbeitung zu bestimmen.

Phase 3. Die Datenvorbereitung gliedert sich in die Schritte Selektion und Integration, Säuberung, Reduktion und Transformation von Daten (Chapman et al., 2000). Die Relevanz einer Datenselektion und -integration resultiert aus den verschiedenen Datenbanken und Quellen, aus denen Daten potentiell entstammen. Nach erfolgter Datenselektion sind die Daten in einer konsistenten Datenbasis mit schlüssigen Datensätzen zu vereinheitlichen. Probleme, die bei der Integration auftreten können, sind unter anderem Entitäten-Identifikationsprobleme, Redundanzen, Widersprüche oder Datenwertkonflikte (Cleve & Lämmel, 2016).

Danach ist der Datenbestand manuell zu bereinigen. Es ist darauf zu achten, dass eingefügte Werte durch Bereinigung informationsneutral sind, ohne eine Verfälschung der vorhandenen Dateninformationen herbeizuführen. Neben fehlenden Daten stellen ebenso verrauschte Daten und Ausreißer oder inkonsistente und falsche Daten mögliche Problemstellung dar, die während des Säuberungsprozesses zu unterbinden sind (Cleve & Lämmel, 2016).

Eine Reduktion der Daten ist dann notwendig, wenn ein Datensatz zur Ausführung des Data Mining zu groß ist. Einerseits kann die Komplexität eines Datensatzes mit Hilfe einer zeilen- oder spaltenweisen Aggregation (Cleve & Lämmel, 2016) verringert werden. Mehrere Daten werden also auf Basis von charakteristischen Attributen zusammengefasst. Ein konkreter Anwendungsfall ist etwa das Clustern von Datenelementen und -kategorien durch Anwendung der spaltenweisen Aggregation. Personenbezogene Daten können so von dem Rest des Datenbestandes abgespalten und kategorisiert werden. Eine zweite Lösung bietet die Dimensionsreduktion als Vorwärtsauswahl oder Rückwärtseliminierung (Cleve & Lämmel, 2016), indem Stichproben einer repräsentativen Teilmenge der selektierten Daten durchgeführt werden. Eine Erfassung aller personenbezogener

Daten kann etwa mittels der Vorauswahl erfolgen, indem alle Daten, die keinen direkten Personenbezug aufweisen, durch sukzessive Aufnahme neuer Anforderungen gelöscht werden.

Ziel der Datentransformation ist die Überführung der Daten in eine brauchbare Form, um in das Data-Mining-System integriert werden zu können. Verfahren zur Datentransformation sind etwa Codierungen, Zeichenketten (z.B. Umlaute), Maßeinheiten und Skalierungen, Kombinationen oder Separierungen von Attributen, Berechnungen abgeleiteter Werte, Aggregationen oder Datenglättungen (z.B. Regression) (Cleve & Lämmel, 2016).

Phase 4. Die Modellbildung des Data Mining unterscheidet zwischen Potential- und Beschreibungsaufgaben. Potentialaufgaben umfassen die Datenklassifikation und das Ableiten von Prognosen, wohingegen Beschreibungsaufgaben der Segmentierung oder dem Aufstellen von Assoziationen zwischen Datensätzen dienen (Cleve & Lämmel, 2016). Bei der Klassifikation erfolgt eine Zuordnung eines Datenobjekts zu einer vordefinierten Klasse entlang charakteristischer Merkmale. Die Prognose hingegen zielt auf die Entwicklung eines Bewertungsmodells zur fortlaufenden Ermittlung stetiger Werte ab. Bisher unbekannte, numerische Merkmale werden auf Basis anderer Merkmale oder erlangter Erkenntnisse vorausgesagt und Abhängigkeiten zwischen diversen Variablen hergestellt. Im Rahmen der Segmentierung wird die Gesamtheit aller Daten in Teilmengen unterteilt und mehrere Datenobjekte mit gemeinsamen Merkmalen zu einer homogenen Gruppe zusammengeführt. Im Fokus der Assoziation steht die Ermittlung und Beschreibung von Mustern zwischen Datenobjekten, die in einer bestimmten Relation zueinanderstehen. Beispiele für Data-Mining-Verfahren sind Entscheidungsbäume, Cluster-Algorithmen oder Regressionen.

Für den vorliegenden Anwendungsfall muss primär eine ganzheitliche Erfassung und Strukturierung aller personenbezogenen Daten vorgenommen werden. Verschiedene Datenkategorien sind zur einheitlichen Dokumentation im VVT zu einem einzigen Datenelement zu reduzieren. Dafür eignen sich die Klassifikation und Segmentierung.

Zu Beginn erlaubt die Klassifikation eine Kategorisierung personenbezogener Daten gemäß charakteristischen Merkmalen, durch die natürliche Personen eindeutig identifizierbar sind. Um eine solche Separierung zu erreichen, müssen dem Data-Mining-System die Merkmalanforderungen bekannt sein. Die Anforderungen ergeben sich vorrangig aus der Definition personenbezogener Daten des Artikel 4 DSGVO. Beispiele für charakteristische

Merkmalanforderungen zur Datenklassifikation sind bspw. Name, Anschrift, E-Mail-Adresse oder Telefonnummer.

Überdies sind die als relevant klassifizierten Datenobjekte zu segmentieren und bestehende Datenkategorien bestimmten Datenelementen zuzuweisen. Eine Möglichkeit stellt das Rule-based Reasoning (Chowdhary, 2020) dar, indem Regeln entlang des Wenn-Dann-Sonst-Prinzips (Frye et al., 1995) erarbeitet werden. Ein Beispiel zur Segmentierung personenbezogener Daten kann etwa über die Regel *„Wenn die Datenkategorie die Angabe Name oder Anschrift enthält, dann sind diese personenbezogenen Daten dem Element persönliche Kontaktinformationen zuzuweisen“* erfolgen. Dieses Schema muss für alle identifizierten Datenkategorien und -elemente umgesetzt werden – unter der Prämisse, dass die Möglichkeit einer Segmentierung besteht. Zur besseren Veranschaulichung ist die Vorgehensweise der Regelbildung in Abbildung 1 dargestellt.

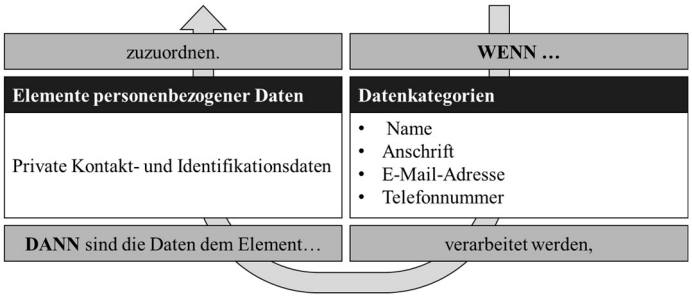

Figure 1: Wenn-Dann-Regel zur Segmentierung personenbezogener Daten.

Hinzu kommt die Notwendigkeit, auf Grundlage der identifizierten und segmentierten Daten resultierende Verarbeitungsprozesse zu erfassen, zu dokumentieren und zu aktualisieren. Hierfür kann auf die Assoziation und Prognose zurückgegriffen werden.

Mit Hilfe des Assoziationsverfahrens lassen sich Abhängigkeiten und Muster zwischen Daten und Verarbeitungstätigkeiten feststellen. Das stellt sicher, dass personenbezogene Daten, die nicht als solche kenntlich sind, identifiziert werden können. Eine Möglichkeit stellt dabei die semantische Interoperabilität (Gödert, 2010) zur Kollaboration zwischen diversen IT-Systemen mittels Klassifikationssystemen, Taxonomien oder Nomenklaturen dar. Die semantische Interoperabilität eignet sich im Speziellen zur Erfassung und Dokumentation von Verarbeitungstätigkeiten auf Grundlage der verarbeiteten personenbezogenen Daten. Das Data-Mining-System wird dazu befähigt, Informationen mit den IT-Systemen der Organisation auszutauschen und so bspw. den Zweck einer Verarbeitung zu ermitteln. Mittels eines intelligenten und vernetzten Zusammenspiels zwischen IT-Systemen lassen sich die zentral im VVT

festzuhaltenden Inhalte und Informationen detektieren und einheitlich dokumentieren. Dadurch wird ebenso eine Standardisierung der Dokumentation erreicht.

Als letzter Schritt unterstützt die Prognose dabei, Zusammenhänge zwischen bekannten und bisher unbekannten Merkmalattributen herzustellen und Trendentwicklungen zu prognostizieren. Treten etwa Änderungen in den Angaben personenbezogener Daten auf, können diese analysiert und bei dokumentierten Verarbeitungstätigkeiten aktualisiert werden. Weiterhin können Risiken aufgrund von Datenlecks oder bei Nicht-Einhalten der geltenden Datenschutzvorgaben präventiv gemeldet werden. Es lässt sich festhalten, dass eine Kombination der vier Data-Mining-Vorgehen essenziell ist, um den geschilderten Herausforderungen und Problemstellungen zu begegnen und die identifizierten Anforderungen zu erreichen.

Phase 5. Im Kontext der Evaluation werden die Analyseergebnisse geprüft. Ob die Umsetzung erfolgreich ist, ergibt sich aus einer Ermittlung des Erfüllungsgrads der initial spezifizierten Erfolgs- und Zielkriterien. Zentrale Fragestellung ist, ob der erwünschte betriebswirtschaftliche Nutzen durch das Entwicklungsergebnis erzielt wird (Cleve & Lämmel, 2016).

Im weiteren Verlauf ist eine Analyse der auftretenden Fehler durchzuführen, woraus sich unter Umständen weitere Optimierungspotentiale ergeben (Chapman et al., 2000). Tritt dieser Umstand auf, kann in eine der vorangegangenen Phasen zurückgekehrt und so das Data-Mining-Vorhaben sukzessive verbessert werden.

Phase 6. Den Abschluss bildet die praktische Implementierung des Data-Mining-Systems. Zur optimalen Einsatzvorbereitung wird ein im Detail ausgearbeitetes und strukturiertes Vorgehen zum künftigen Monitoring des Data Mining und der resultierenden Analyseergebnisse vorgesetzt. Des Weiteren muss eine ausreichende Motivation der Mitarbeitenden der Organisation, in der das System Anwendung finden soll sowie eine umfassende Unterstützung der durch das Data Mining betroffenen Mitarbeitenden (z.B. IT-Abteilung, Datenschutzbeauftragter, Fachabteilung etc.) gegeben sein, um das Scheitern des Projekts zu verhindern. Das System ist in den Regelbetrieb der Organisation zu überführen und in laufende Prozesse einzubetten.

Diskussion

Die ersten drei CRISP-DM-Phasen beanspruchen etwa 50 bis 70 Prozent des Arbeitsaufwands zur Entwicklung des Data Mining, wobei die einzelnen Phasen manuell vorzubereiten und umzusetzen sind (Wuttke, 2020). Ein direkter Vergleich des Status Quo zur Pflege eines VVT und den Phasen

des CRISP-DM impliziert, dass die initialen drei CRISP-Phasen gleichermaßen im manuellen Pflegeprozess eines VVT stattfinden. Unternehmen, die bereits ein VVT pflegen, haben die Schritte im Optimalfall durchlaufen. Auch wenn zur Pflege eines VVT kein Data-Mining-System etabliert werden soll, ist es sinnvoll, die Phasen gewissenhaft umzusetzen. Organisationen sollten nach Ausführung der initialen Phasen in Erwägung ziehen, ihre Ergebnisse in KI-Algorithmen und Regeln zu überführen und die manuellen Arbeitsaufwände auf ein Data-Mining-System zu verlagern. Unternehmen können auf ihrem bisherigen Arbeitsstand aufbauen, die Inhalte entsprechend dem dargestellten Vorgehen anpassen und letztlich in ein Data-Mining-Modell überführen. Jedoch muss das CRISP-DM-Modell nicht zwangsläufig in der Implementierung eines Data-Mining-Systems münden, auch wenn dies zu einem deutlichen Anstieg der Produktivität beiträgt. Stattdessen sehen sich Organisationen aufgrund der Rechenschaftspflicht ohnehin damit konfrontiert, Transparenz hinsichtlich der Datenverarbeitungen und -flüsse sicherzustellen. Um ebendiese Transparenz zu erreichen, haben Organisationen die initialen Phasen Business Understanding, Data Understanding und Data Preparation zur Erfassung, Dokumentation und Aktualisierung aller existierenden Verarbeitungsprozesse aufzubereiten. Folglich zieht das dargestellte Vorgehen kein Mehraufwand nach sich, sondern bietet Organisationen im Gegenteil die Möglichkeit, Synergien zu nutzen und in Zukunft bedarfsorientiert auf ihrem bisherigen Arbeitsstand aufzubauen, um eine technologische Unterstützung und Optimierung des Vorgehens zur Pflege eines VVT herbeizuführen.

Zwar kann Data Mining bei der vollständigen Erfassung und Kategorisierung personenbezogener Daten, dem Ableiten von Verarbeitungstätigkeiten sowie der Dokumentation und Aktualisierung im VVT unterstützten. Jedoch sind nachgelagert weiterhin manuelle Aufwände notwendig. Bspw. verantwortet das Rechtswesen einer Organisation die Zentralisierung von Verarbeitungstätigkeiten im VVT, die Überführung neuer datenschutzrechtlicher Vorgaben in konkrete Anforderungen an Data Mining oder die konstante Überwachung der Qualität der Analyseergebnisse.

Eine weitere Handlungsempfehlung betrifft die Erweiterung des CRISP-DM-Modells um die Monitoring-Phase. Neben einer fortwährenden Wartung des Systems sind auch die Ergebnisse der Datenanalysen durch Verantwortliche der Datenschutzorganisation (z.B. Datenschutzbeauftragter etc.) zu überwachen, da diese bestens mit den rechtlichen Grundlagen vertraut sind. Es ist sicherzustellen, dass alle Mitarbeitenden, die in Zukunft Berührungspunkte mit

dem System haben, umfassend geschult werden. Darüber hinaus ist ein Datensicherheitskonzept gemäß Artikel 32 DSGVO zu entwickeln, um softwareseitige Störungen und Systemausfällen vorzubeugen. Im Falle eines Absturzes wird etwa der Zugriff auf das zentrale Verzeichnis verweigert. Daher müssen in regelmäßigen Abständen automatische Backups des VVT durchgeführt werden und Mitarbeitende der Organisation dafür Sorge tragen, das System durch Schutzmaßnahmen (z.B. technisch-organisatorische-Maßnahmen) abzusichern.

Data Mining erlaubt eine ganzheitliche, transparente und zentrale Steuerung des VVT-Prozesses, sodass ein technologischer Einsatz zur Wahrung der Konformität geeignet ist. Durch Einhalten der Vorgaben des Artikel 30 DSGVO wird das Haftungsrisiko wesentlich reduziert und der aktuell gelebte Prozess in der Praxis flexibilisiert und vereinfacht. Die ersten vier Phasen des Modells können als Basis für vergleichbare Aufgabenstellungen genutzt und spezifiziert werden.

Nichtsdestominder zeigt sich der Unterstützungsgrad des Data Mining erst durch Entwicklung entlang einer realen Aufgabenstellung und tatsächlichen Implementierung des Systems. Darüber hinaus ist die Modellentwicklung auf Annahmen und Theorien gestützt. Die CRISP-DM-Phasen wurden zwar detailliert ausgearbeitet, jedoch ohne die technologische Ebene vertieft zu betrachten. Dies ist unter anderem der Tatsache geschuldet, dass kein Testmodell unter realen Umständen entwickelt und zu Testzwecken implementiert wurden. Eine Aussage über die tatsächliche Um- und Einsetzbarkeit des Data-Mining-Systems kann somit nicht getroffen werden. Grundsätzlich ist die technische Sicht bei der Entwicklung des Vorgehensmodells unterrepräsentiert. Es kann vorkommen, dass Probleme, die in der Anwendung und Programmierung des Data-Mining-Systems auftreten, nicht vollständig erkannt und berücksichtigt wurden. Daher wird empfohlen, das System in der Praxis zunächst umfassend zu testen und mit erfolgreichem Abschluss der Testphase auf weitere Bereiche des Unternehmens auszuweiten. Ferner können in Zukunft vor- und nachgelagerte Teilprozesse oder weitere Anforderungen der DSGVO berücksichtigt und auf ähnliche Weise intelligent gesteuert und optimiert werden.

Es bleibt zu erwähnen, dass das CRISP-DM-Modell einem Standardmodell entspricht, welches der Standardisierung diverser Anwendungsfälle dient und damit nicht nur auf die Pflege eines VVT begrenzt ist. Stattdessen kann das CRISP-DM-Modell beliebig erweitert und ebenso auf andere Ausgangssituationen übertragen werden. Das Modell kann demnach genau wie VVT in der Praxis vielfältig ausfallen. Aus diesem Grund findet im Rahmen des vorliegenden Beitrags eine

generische Darstellung und Entwicklung des Modells Anwendung. Auf die Beschreibung einzelner, konkreter Anwendungsfälle wird in diesem Zusammenhang bewusst verzichtet.

Als weiterer Forschungsbedarf kann das Process Mining und dessen Kombination mit Data Mining betrachtet werden. Process Mining vereint die Vorteile des Data Mining mit denen der Prozessmodellierung, sodass eine effiziente Überwachung und Erstellung von komplexen Echtzeitprozessen möglich ist (Reinkemeyer, 2020). Durch Standardisierung von Prozessen kann einerseits die Transparenz erhöht werden und damit Schwachstellen der aktuellen Prozessumsetzung effizient geprüft und bei Bedarf verbessert werden (Peters & Nauroth, 2019). Andererseits werden durch Prozessautomatisierungen Redundanzen reduziert, Engpässe vermieden und damit einhergehend Kosten reduziert werden (Peters & Nauroth, 2019). In einem nächsten Schritt kann das System um maschinelles Lernen erweitert werden. Diese Form der künstlichen Intelligenz ermöglicht die Entwicklung von Handlungsempfehlungen und Generierung von Maßnahmen anhand großer Datenbestände eines Prozesses (Reinkemeyer, 2020). Gemeinhin wird dieses Vorgehen als Predictive Process Mining bezeichnet. Anhand prozessbezogener Daten erkennt das System relevante Kausalitäten und erklärt diese. Während das System automatisch Trends und Muster ableitet, werden die entwickelten Maßnahmen durch Mitarbeitende des Unternehmens bewertet und schließlich umgesetzt.

In einer vertieften Betrachtung sind die Potentiale einer Kombination des Process- und Data Mining zur Erreichung einer ganzheitlichen Standardisierung und Sicherstellung datenschutzrechtlicher Vorgaben zu untersuchen.

3. The Magic Triangle of Data Governance - Multidimensional risk assessments for data governance programs

Published: Data Governance Insights – The Magic Triangle of Data Governance, Deloitte GmbH (forthcoming 2023)

Authors: Wolfgang Köhler, Christian Schultz, Christoph Rasche, Andreas Herzig

Introduction

Data protection, privacy regulations, and legislations are diverse and changeable. Due to the increased use of technologies, the growing importance of data ensures dynamic development and further progresses rules and laws across industries and countries. Legislators are trying to keep pace with rapid technological advancements. Apart from the fact that legislation is diverse, sometimes sector-specific, constantly changing, and sometimes incompatible, there is a problem of definition and application of many standard legal terms and concepts. Examples are ownership, authorship, or the financial assessment of an amount of damage when we talk about data, data processing, and infringements. For companies, some focus topics are apparent. It is about efficient processes through better algorithms, higher computing capacity, larger amounts of data, and associated opportunities for increasing process automation. Besides, the use of IOTs in the Smart Factory, Smart Cars, Smart X world, and the generation of partly unforeseen data-driven insights through new analysis and source combination methods are more than ever in the spotlight.

Bringing together regulatory requirements on the one hand and technical developments in business, on the other hand, is a key challenge today. Data governance and compliance require the analysis and implementation of a wide range of business areas and processes. For companies, complexities arise in many respects from the diversity and continuous change of data regulations. The business focus and the business partners involved are rarely limited to one single market and, therefore, only single relevant legislation. In addition to these challenges, which can still be planned or illustrated, there is also the fact that the customers are not limited to specific areas, services, or products anymore. People cross borders and different legal areas with possibly contradictory regulations while using smart devices or connected services, and data processing and transmission are ongoing. Such use cases may have consequences for the data controller and processors and require transparency about international laws, applying those laws, and their differences.

The question of which data regulation perspectives must be considered to meet data governance requirements will be addressed. This article describes the categorization of international data regulations that need to be considered and their relevance to achieve and validate data governance.

The individual as the subject of protection

With a view to international data regulations, more comprehensive privacy regulations seem apparent, as famous, fundamental, and controversially discussed innovations have come into effect in recent years. The EU GDPR, the California Consumer Privacy Act (CCPA), and the Brazilian General Data Protection Law (LGPD) are prominent examples of privacy laws and bills worldwide that mainly focus on protecting natural persons' rights and freedoms. Their focus is on individuals and the protection of personal rights, primarily informational self-determination and the right to privacy, and the implementation of security and transparency obligations for the processors of personal data. When raising national or regional requirements to the international level, cross-border data transfer is of fundamental importance. Single laws can affect different entities even if the primary focus is on one.

A large number of different data regulations alone implies an enormous complexity. The idea that compliant data processing and monetization can only be achieved if an internationally operating company respects and considers all different privacy laws and follows new developments as closely as the publication of new local regulations. How should the management of an international company identify the "right" approach to implementing the laws? After all, as much data as possible should be processed centrally to profit fully from underlying potential, regardless of its local origin. That contributes to the development of international digital products; the quantity and quality of available data is an asset in multiple areas today. The need for centralized data strategies approaches, and data compliance frameworks become apparent when comparing practicability/feasibility and the internal effort required to implement regulations. Hardly anyone would like to look at each national legislation individually and start from scratch to implement each legal innovation - primarily when similar requirements have already been successfully implemented elsewhere. Besides, there is room for interpretation, which new laws without similar case law/court rulings can show. Moreover, it happens that local peculiarities contradict each other in an international context. Which law predominates another, or could some even violate the right to privacy? Can managers even decide this without acting incompliantly or committing a legal violation?

Business & markets as the subject of protection

Suppose you feel content with the knowledge such as "our enormous effort in implementing privacy regulations in the past has paid off [...] we can achieve all this with our framework". In that case, it is essential to focus on further data topic regulation. Of course, the right to privacy must be a fundamental element of any business, but it must not result in other data regulations being neglected.

Similar to privacy regulations, some data regulations focus on protecting businesses and markets. For example, the EU Directive on the protection of trade secrets, the EU directive on copyright in the digital single market, and the EU regulation on the free flow of non-personal data are all laws that focus on protecting and supporting businesses or markets. Furthermore, data localization obligations, which require the local storage and processing of data and the operation of servers and data centers in the respective countries, should be mentioned.

One well-known example is the Russian Data Localization Law. When collecting personal data, including through the information and telecommunication network, an operator must document the recording, systematization, accumulation, storage, adjustment (update or alteration), and retrieval of the personal data of citizens of the Russian Federation using databases located in the territory of the Russian Federation (see Article 2, paragraph 1 of Russian Federal Law No. 242-FZ). Additionally, parts of the previously mentioned laws serve economic protectionism and national companies' interests. International data regulations with business as the subject of protection are also diverse and continuously changing. The complexity in this area is also considerable. Another example is patent law, which in many jurisprudences forms a separate sub-area of private law. Today it is still extraordinarily dynamic and complex, so that there are many lawsuits, although the field of law is not new. The number of patents that, e.g., a modern vehicle could infringe, and with it, the number of patent holders who could claim royalties, has multiplied over the past decade. Vehicle manufacturers were exposed to numerous patent lawsuits with unknown or irrelevant constellations a few years ago. The plaintiffs and their patents range from chip technology to mobile phone standards such as LTE. Based on the potential financial gain alone, patent holders might decide to bring their claims to the OEM and not one of the smaller suppliers. The questions that are arising: How can management, besides the privacy regulations, implement the business-relevant data regulations in an international context? How can the management identify the relevant regulations for its own business? How to implement these

globally - especially when the company's knowledge and possibly a competitive advantage is at stake? How does management best deal with local specifics and with inconsistencies or contradictions in individual laws? Where are similarities and differences with the data regulations of the other subjects of protection?

The public as the subject of protection

In addition to the individual privacy and business-focused regulations, some data regulations focus on national security or protection of the state and the public. Different motivations may justify these regulations. On the one hand, it concerns the area of national security, such as the protection of important organizations, as in the energy and telecommunications sectors, whose undisturbed operations can be a direct factor in national security, and in an international context, the restriction, and control of access by foreign states to specific information. On the other hand, data and information must ensure continuous compliance monitoring with national legislation and enable law enforcement.

In particular, the Clarifying Lawful Overseas Use of Data Act (CLOUD Act) and parts of the Cybersecurity Law of the People's Republic of China ('CSL') should be emphasized here. The CLOUD Act allows US federal law enforcement agencies to compel US-based technology companies, by warrant or subpoena, to provide data stored on servers upon request, whether the data are stored in the US or on foreign soil. The CSL requires critical information infrastructure operators ('CIIOs') to store personal information and essential data generated from China's critical information infrastructures. The legal requirements of some state and public-focused regulations are stringent on cross-border data transfer. For example, the Russian Law on Personal Data (Federal Law No. 242-FZ) stipulates that the storage and updating of data on Russian citizens are limited to the resources of data centers within the Russian Federation, better known as data localization. Regulations from this realm may be particularly contradictory to national privacy laws that seek to protect natural persons' rights and freedoms. The contradictions mentioned above exist, for example, concerning third parties' access to data or the transfer of data to third parties (e.g., various authorities).

The crucial questions are, which requirements affect the business at all, in what form, and to what extent? Do they result in necessary measures about data governance? How can management, besides the privacy and business-focused regulations, consider the national security and public-focused data regulations in an international context? What needs to be considered by managers

when developing the European, Russian, Chinese, or American (etc.) markets? Are datacenters needed in each country, and can the locally available data be used further, and to what extent? Are there legitimate privacy concerns when choosing a partner in the USA, China, etc., to process personal data?

Using the magic triangle of data governance

Management decisions often require making predictions, conducting intensive analyses, and weighing opportunities and risks. The three perspectives described above can help make a business decision and review the current company's implementation of data regulations in an international context and its data governance maturity.

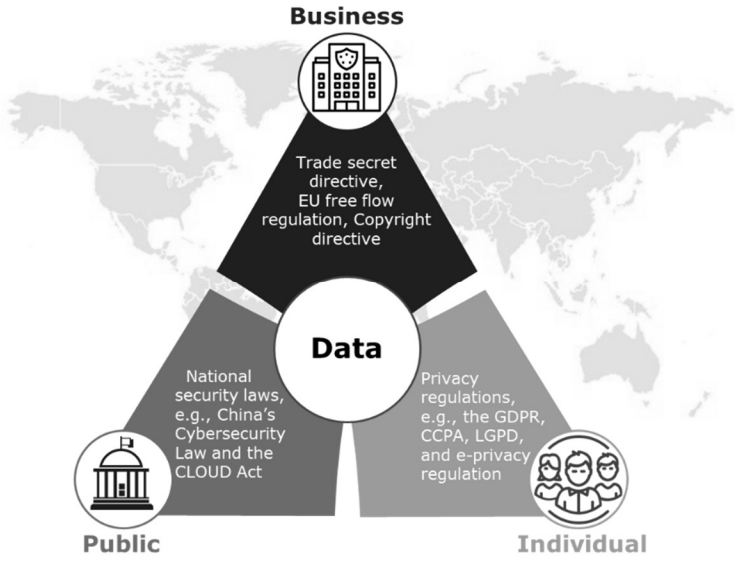

Figure 2: The Magic Triangle of Data Governance (Köhler et al., 2022).

Single use-cases are an excellent place to start and prevent businesses from being overwhelmed. Both current and future scenarios can be described and analyzed and documented using the system described below. Use cases play a decisive role and should be developed and enhanced parallel to product and market development. Identified challenges can thus be considered and implemented as early as possible.

Before starting the first step of analyzing the situation, it seems logical to answer three main questions and consider the magic triangle, three protection subjects.

a) Who and where is my customer?

b) Where do I run my business?

c) Where do my business partners and providers run their business?

The result is a multitude of possible constellations that should be mapped in the use cases. In this context, the underlying international data flows play a significant role. On the one hand, the different locations of their own business, the service providers, and communication between them. On the other hand, and just as important is the question of who and where my customer is. Especially in the connected services area, the customer is a moving target. Customers can cross or enter different legal jurisdictions while data processing and transmission are in progress. The demand for the use cases to be analyzed is to map precisely these dynamics.

Use case assessments

A Chinese car manufacturer with international partners offers vehicles with various digital components and connected services in Europe. Data is continuously generated, transmitted, and processed by the manufacturer as well as the digital service provider in the US. The European customer enjoys many connected services while driving the car.

To assess the use case for customers in Europe, this means:

a) Who and where is my customer? (e.g., European customer)

1. What data legislations must be observed to protect the rights and freedoms of the natural persons affected in Europe following the law.

2. Are there any business-related laws or regulations at the manufacturer or one of its business partners that could have a negative impact on individuals' fundamental rights and freedoms in Europe?

3. Are there any data regulations that focus on national security or protection of the state and the public at the manufacturer or one of its business partners that might affect individuals' fundamental rights and freedoms in Europe?

To assess the use case with the focus on the vehicle manufacturer's perspective, this means:

b) Where do I run my business?

1. What data requirements must be observed in China to protect natural persons' rights and freedoms following the law when processing data? Which laws and standards for the protection of personal rights are to be implemented in the various markets?

2. What are the relevant and most important data regulations to protect and support China's own business? How can these be implemented globally, especially when its knowledge and possibly a competitive advantage are at stake?

3. What data regulations that focus on national security or protection of the state and the public in China affect the business? How and to what extent? Do they result in necessary measures concerning data governance?

To assess the use case with a business partner's perspective, this means:

c) Where do the business partners/ service providers run their business? (e.g., Business Partner located in the US)

1. What data requirements must be observed in the US to protect natural persons' rights and freedoms following the law when sharing data with business partners? Which rules and standards for the protection of personal rights are to be implemented in the various markets?

2. What are the relevant and most important data regulations in cooperating with partners to protect and support the US's own business? How can these be implemented globally, especially when its knowledge and possibly a competitive advantage are at stake?

3. What data regulations focus on national security or protection of the state and the public in the US affect the business? How and to what extent? Do they result in necessary measures concerning data governance?

Assessment in the use cases context

The first step is taken once the three perspectives of individual, public, and business have been analyzed and documented for the relevant legal data. There is transparency about the regulatory circumstances to be taken into account - for each of the areas of application. Afterward, as in the operative business, it is essential to put them into context with each other to identify and analyze the real operational context problems and derive individual action steps or a strategy.

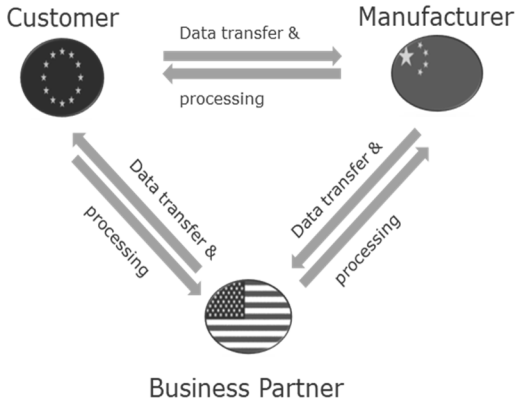

Figure 3: Magic Triangle Context Assessment.

The magic triangles consider all three angles of data-related regulations and laws for all stakeholders and their relevant jurisdictions. By rotating the individual aspects about which transparency has already been established, different requirements per jurisdiction and subject of protection are comparable. That allows existing challenges and inconsistencies in individual laws to be identified, evaluated, and management decisions derived from them. Since some of the associated problems will continue to exist in the future and for other use cases, management can develop a strategy and decide how these challenges will be consistently addressed.

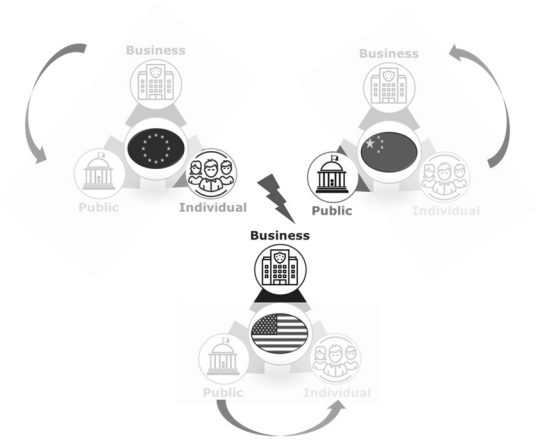

Figure 4: Magic Triangle Rotation.

Critical Questions of the selected use case

European customer
Subject of protection: Individual

- What data are processed on what legal grounds?
- Where to store & process the data?
- How to transfer data to China and to the US service provider?
- How to implement data subject rights?
- Are there differences to be considered depending on the location of the customer?
- Is it permitted to process the customer's location?

Chinese manufacturer
Subject of protection: Public

- What data need to be stored in China?
- How to transfer data to Europe, how to the US?
- What regulatory oversight will apply?
- What are the effects of state powers of supervision, investigation, and enforcement?
- How to implement data subject rights?

US digital service provider
Subject of protection: Business

- What information must the service provider disclose?
- What are the requirements for the programs and algorithms used?
- Is the intellectual property of the manufacturer and the provider sufficiently protected?
- Are there standards and norms for programs used?
- Who is liable for which cases?

4. Developing digital products with compliance-driven personal data integration

Abstract

The digital era has revolutionized the development of products and services, transforming business operations and competition. Organizations leveraging digital technologies to create novel products and services must adapt to changing customer preferences and capitalize on data-driven innovations. However, integrating personal data into the development process presents challenges in terms of data privacy and transparency. This submission outlines pragmatic strategies for personal data integration in digital products, emphasizing privacy principles, regulatory compliance, and trust. By prioritizing transparent and responsible data practices, these strategies balance data collection benefits and privacy concerns, ensuring user transparency and control. Informed by legal mandates, decision heuristics support this approach. Proactively adopting responsible data practices enables companies to develop innovative, user-centric digital products that effectively meet user needs while protecting privacy.

Keywords: Data Privacy; Digital Product Development; Transparency; Purpose Limitation; Consent; Legitimate Interest; GDPR

Published: The XXXIV ISPIM Innovation Conference, Ljubljana, Slovenia on 04 June to 07 June 2023. Editors: Iain Bitran, Leandro Bitetti, Steffen Conn, Jessica Fishburn, Paavo Ritala, Marko Torkkeli & Jialei Yang. ISBN 978-952-65069-3-*7.*

Authors: Wolfgang Köhler

Introduction

The digital era has seen a remarkable rise in the development of digital products and services, resulting in a significant shift in how businesses operate and compete by increasingly replacing their physical counterparts (Loebbecke and Picot 2015). Driven by digital technology-based innovations, such as digital business models, platform innovations, product innovations, and marketing innovations (Varadarajan et al. 2022), this transformation has given birth to highly successful digital native firms and redefined legacy firms globally.

Organizations leveraging digital technologies to create new products, services, and business models (Legner et al. 2017; Nambisan et al. 2017) must adapt to novel customer preferences and behaviors (Lyytinen, Yoo and Boland 2016). Furthermore, digital product and service innovations often initiate follow-up innovations in the form of complementary services or products (Fichman, Dos Santos and Zheng 2014), expanding their impact on the marketplace. By utilizing cutting-edge technologies, best practices, and cross-functional collaboration, companies can develop user-friendly and innovative digital products that cater to user needs (Liao, Chen and Yen 2007). This focus on understanding and fulfilling user expectations is crucial for fostering positive attitudes, satisfaction, and continued usage of digital products and services, ultimately improving product quality and user satisfaction.

In addition, the increasing availability and accessibility of customer data enable data-driven innovations (Akter and Wamba 2016; Dinter and Krämer 2018; Willing, Brandt and Neumann 2017), presenting opportunities and challenges concerning data transparency, privacy, and security (Spiekermann et al. 2015). When personal data is involved, digital product development and new technologies introduce new privacy protection challenges, such as a lack of control and transparency, data reusability, data inference, re-identification, and profiling (ENISA, 2021). Data privacy is a crucial ethical issue as the collection and utilization of personal data grow, requiring the protection of individuals' autonomy, dignity, and confidentiality while using their data in ways that align with their interests and expectations (Altman 1975; Nissenbaum 2009; Westin 1968). Businesses must adhere to privacy regulations such as the European Union's General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and Japan's Act on the Protection of Personal Information (APPI) and ensure transparency and individual control over personal data. During digital product development, companies must establish and maintain transparency for affected individuals, select the appropriate legal basis, and obtain valid consent to

address privacy core principles while optimally and compliantly using available and new data (Liu, Pavlou and Cheng 2022; Wieringa et al. 2021). Balancing the benefits of data collection with individual privacy concerns requires adopting a forward-looking approach to responsible data practices, even as new processing purposes emerge during development.

This submission proposes practical strategies for integrating personal data in digital product development while ensuring legal compliance and trustworthiness. It focuses on GDPR Article 5's fundamental privacy principles, such as lawfulness, fairness, transparency, and purpose limitation, as the foundation for data processing (Figure 5) and prioritizes data subjects' empowerment with greater control. By addressing the significant challenges of regulatory compliance and trust when using personal data in digital product development, this approach emphasizes responsible data practices rather than engaging in an ethical discussion on personal data usage. We derive decision-making heuristics from relevant legal frameworks, enabling legally compliant development and enhancement of personal data-based digital products. Additionally, we examine various legal bases from privacy professionals' perspectives.

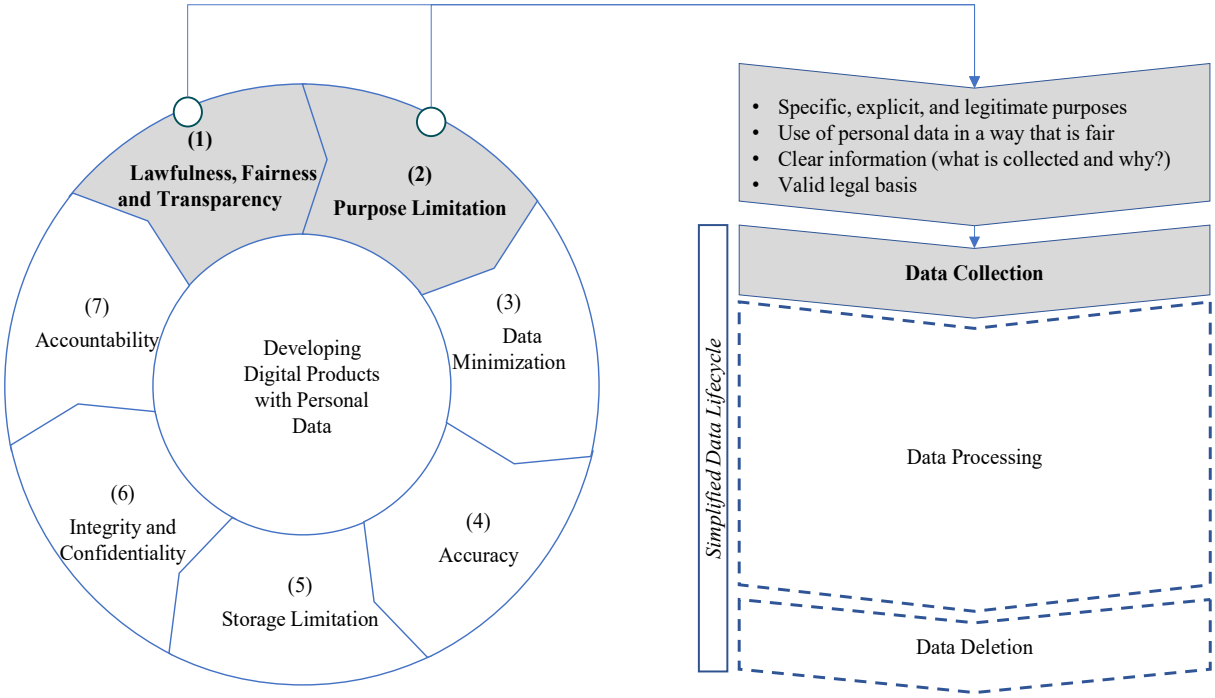

Figure 5: Lawfulness, fairness, transparency, and purpose limitation, as the foundation for personal data processing.

Developing Digital Products utilizing Personal Data

Digital products, as delineated by Lyytinen, Yoo, and Boland (2016), comprise goods or services either embodied in or enabled by information and communication technologies. Hui and Chau (2002) characterize them as digitizable items, including software, music, and various publications, which are increasingly accessible and marketable online. Wang, Wang, and Yao (2005) emphasize that digital products can be transacted and delivered through the internet, integrating information, payment, and delivery into a single online channel. Zhang and Jiang (2001) highlight consumption characteristics such as non-exclusiveness and non-rivalry. These products have unique attributes, including attrition-free, changeable, and replicable properties with low marginal manufacturing costs (Shapiro and Varian 1998). Hui and Chau (2002) categorize digital products based on trialability, granularity, and downloadability into utilities and tools, content-based digital products, and online services.

We define digital products as goods or services, either embodied in or enabled by information and communication technologies, that can be digitized, transacted, and delivered via the internet. These products exhibit attributes including being attrition-free, changeable, replicable, and low marginal production costs.

The rapid innovation driven by digital technology presents challenges in controlling and predicting development processes (Henfridsson, Mathiassen and Svahn 2014; Yoo et al. 2012; Yoo et al. 2010). Cooper and Kleinschmidt (1993) emphasized that customer-oriented enterprises must be fully aware of customer needs, market competition, and market nature during new product development. These factors are critical to success.

Personal data can enhance digital products by offering insights into user behavior and preferences, allowing developers to tailor products and improve personalization (Anshari et al. 2019; Zhan et al. 2018). Personalized digital products have been demonstrated to boost user satisfaction, loyalty, and purchase intention, leading to increased revenue and market share (Wong 2012; Bauer and Leker 2013; Wamba et al. 2015). As technology advances and generates vast amounts of user data, firms that leverage this information to develop new features or products are more likely to succeed (Sarin and O'Connor 2009; Roberts and Candi 2014). Firms must offer a clear value proposition to incentivize users to share their data while ensuring responsible and transparent data collection and use (Porter and Heppelmann 2014). A collaborative approach and focus on user needs and preferences are essential for developing digital products. By responsibly

and transparently utilizing personal data, organizations can create innovative and personalized products that satisfy customers, ensure user privacy, and comply with privacy regulations, ultimately reducing compliance risks and driving business growth.

Big Data, Big Challenges

In the digital transformation era, product development has evolved towards creating innovative digital solutions catering to a technology-driven and interconnected world (Berman 2012). The growing prominence of Big Data in new product development efforts has led global firms to recognize the competitive edge gained through valuable insights from vast data sources (Zhan et al. 2018; Barton and Court 2012; Salehan and Kim 2016). Artificial intelligence (AI) has emerged as a vital component in innovation, generating real value by reducing risks and costs (Haefner et al. 2021). Consequently, businesses increasingly leverage synergies between Big Data and AI to gain deeper customer understanding, develop improved products, and offer more personalized services (Zhan et al., 2018). However, the rapid growth of Big Data and AI applications has raised privacy and data protection concerns (Forgó, Hänold and Schütze 2017). Ensuring compliance with data protection principles like purpose limitation has become critical to development (Biega and Finck 2021). Navigating challenges posed by purpose limitation and maintaining transparency in AI and Big Data applications is complex (Ghani, Hamid and Udzir 2016; Felzmann et al. 2019).

Purpose Limitation Challenges

Article 5 GDPR mandates that personal data collection and processing be specific, explicit, and legitimate without further incompatible processing. Applying purpose limitation to big data analytics encounters obstacles such as increasing complexity in specifying data collection and processing purposes, potential conflicts with the principle of compatibility, and rapid advancements in data-driven technologies like AI and machine learning (Forgó, Hänold and Schütze 2017; Ghani, Hamid, and Udzir 2016; Hahn 2021). Scholars advocate for nuanced, context-dependent approaches to purpose limitation, striking a balance between big data analytics' potential benefits and respecting individual privacy rights (Forgó, Hänold and Schütze 2017). Biega and Finck (2021) emphasize the importance of purpose limitation in data protection law, while Hahn (2021) suggests viewing it as a dynamic principle adaptable to changing technological, societal, and legal contexts, promoting transparency and accountability in data processing.

Data subjects must be clearly informed of specific processing purposes, with details provided before initiating or continuing processing, mainly when new purposes arise during product development. The data controller is responsible for ensuring and maintaining transparency regarding processing purposes. Ensuring information obligations are met and providing full transparency safeguards the validity of the legal basis. A privacy dashboard, for example, can serve as a digital solution that promotes transparency and data subject control.

Transparency Challenges

Critics argue that the current practice of notice and consent in data protection law is insufficient for genuine informed consent, impacting user trust and acceptance of digital products (Ben-Shahar and Schneider 2014; Solove 2013). Wulf and Seizov (2022) found that GDPR requirements for AI disclosures are vague, leading to heterogeneous and potentially incomplete information. To improve the GDPR, establishing more concrete information requirements regarding AI transparency and explainability and explicitly allowing visualization techniques for consumer information is necessary.

Addressing transparency challenges under the European General Data Protection Regulation (GDPR) requires a multifaceted approach, encompassing both prospective and retrospective transparency (Paal and Pauly 2018; Wachter, Mittelstadt and Floridi 2017). Felzmann et al. (2019) propose a relational approach to transparency, emphasizing trustworthiness as an accountability indicator to address digital product development challenges. Incorporating algorithm audits and the What-If Tool and adopting a dual disclosure strategy can promote transparency in AI systems (Chen, Mislove and Wilson 2015; Sandvig et al. 2014; Venkatadri et al. 2018; Wachter, Mittelstadt and Floridi 2017; Wulf and Seizov 2022). Addressing transparency challenges in Big Data technologies under the GDPR necessitates a nuanced approach balancing privacy, security, and innovation (Wachter, Mittelstadt and Floridi 2017). Collaborative efforts between various stakeholders, including engineers, social scientists, lawyers, philosophers, and ethicists, are essential in overcoming these challenges.

Felzmann et al. (2019) emphasize the necessity for multidisciplinary research to integrate legal transparency requirements into technical systems, including digital product development. Collaborative efforts between engineers, social scientists, lawyers, philosophers, and ethicists can aid in implementing transparency requirements from the design stage. Policymakers should evaluate the usefulness and limitations of the current transparency regime, considering

performative aspects and user constraints (Draper and Turow 2019), and create meeting spaces for policymakers and user-centered researchers to enhance transparency, understanding, and demands in digital product development.

Ensuring Lawful Processing in Digital Product Development

This section focuses on the lawfulness of processing within digital product development. We summarize the requirements for the lawfulness of processing and provide an overview of the legal requirements for consent and legitimate interest. Additionally, we discuss key challenges and considerations when applying these legal bases to digital product development and offer guidance on when each legal basis is appropriate for existing and new data for product development, supported by a decision tree. Additionally, we describe considerations for changing legal bases and present the results of an expert survey regarding the flexibility and permanence of legal bases and their effectiveness in meeting transparency, accountability, security, and data subject control.

Lawfulness of Processing (Art. 6 GDPR)

Article 6 of the GDPR sets requirements for lawful personal data processing, mandating that at least one condition from Paragraph 1 be met. These conditions include (a) obtaining the data subject's consent for specific purposes; (b) processing personal data necessary for contracts or pre-contractual negotiations involving the data subject; (c) fulfilling the data subject's legal obligation; (d) protecting the data subject's or a third person's vital interests; (e) processing data in the public interest or exercising official authority; and (f) processing data based on legitimate interests, unless the data subject's fundamental freedoms prevail. The last condition does not apply to processing operations carried out by public authorities (European Union, 2016). Article 6, Paragraph 4 of the GDPR addresses subsequent personal data processing when the processing purpose changes. The controller must assess the new purpose's compatibility with the original purpose, considering the data nature, potential consequences for data subjects, and safeguards such as pseudonymization. Additionally, the controller must evaluate the relationship between the original and subsequent processing and determine whether the data subject's interests take priority over the controller's or third party's interests (European Union 2016).

Legal obligation refers to personal data processing necessary to comply with legal obligations, like tax or employment laws. Vital interests apply when processing personal data is necessary to protect someone's life or physical integrity. Public interest or official authority applies when public authorities or government bodies perform official duties requiring personal data processing. Performance of a contract applies when processing personal data is necessary for contract performance or pre-contractual steps. Consent and legitimate interest are the primary focus, as they are most relevant for digital product development.

Processing Personal Data Based on Consent (Art. 7 GDPR)

Article 7 of the GDPR specifies the conditions for obtaining and managing consent for processing personal data (European Union 2016). The controller is responsible for demonstrating that the data subject has granted consent for processing their data. Consent must be presented distinctly and clearly, separate from other requests or matters. Data subjects have the right to withdraw their consent at any time without facing negative consequences. Controllers are prohibited from making a service conditional on obtaining consent to process personal data that is not essential for contract performance. Consent must fulfill the following criteria: freely given, specific, informed, unambiguous, and provided through explicit affirmative action (European Union 2016). The consent request should be written in clear, plain language and be easily accessible. Consent can be withdrawn at any time and must be genuinely optional, granted for each specific purpose. General agreements for processing are not considered specific consent. Controllers cannot provide inferior service to individuals who refuse or withdraw consent. Judicial interpretations can help clarify the application of privacy laws, including consent as a legal basis for data processing (Bygrave 2014). The controller must comply with certain conditions and restrictions when processing personal data based on consent. Article 6, Paragraph 4 of the GDPR permits data processing based on consent for purposes "compatible" with the original purposes consented to by the data subject. Compatibility determination requires considering the context of the data collected and the potential consequences of further processing (European Union 2016).

Consent in Digital Product Development

Consent-based personal data processing is crucial in digital product development, as it empowers users to maintain control over their data and fosters trust in the system (Wu et al. 2012).

Research indicates that information transparency significantly impacts privacy concerns and trust, with studies showing that clear disclosure of information handling practices reduces privacy concerns, subsequently improving user trust and willingness to provide personal information (Culnan and Milberg 1998; Laufer and Wolfe 1977; Hinde 1998; Dinev and Hart 2006). Obtaining valid informed consent enhances user trust and confidence, leading to increased adoption and engagement, and helps organizations comply with privacy regulations, thereby mitigating legal risks and reputational damage. Nonetheless, challenges persist in processing personal data based on consent, such as ensuring consent is freely given and specific, providing users with an accessible means to withdraw consent, and ensuring a transparent and user-friendly consent process that meets legal requirements while promoting a positive user experience (Nissenbaum 2004; Schermer 2011).

Valid consent for processing personal data is essential for fostering user trust, ensuring legal compliance, and promoting ethical data practices in digital product development. However, challenges in obtaining valid consent persist due to the privacy paradox, lack of user knowledge, and informational asymmetry between users and service providers (Barnes 2006; Nissenbaum 2009; Taddicken 2014). Organizations can implement simplified language and layered notices to address these challenges to provide concise, easily accessible information about data processing activities (Nissenbaum 2004; Schermer 2011). Granular consent options can empower users by offering more control over their personal information (Goicovici 2019). Just-in-time notices can combat consent fatigue by presenting relevant information when data processing is about to occur (Solove 2013).

Organizations must ensure consent is freely given without duress, deception, or undue influence, by providing genuine alternatives and avoiding negative consequences for withholding consent (European Union 2016). A transparent, user-friendly consent process that meets legal requirements and provides a positive user experience is needed to address the privacy paradox and informational asymmetry. Solutions include restructuring privacy policies, clarifying policy language, implementing standardized policies, and using multilayered policies for increased comprehension (Good et al. 2005; Kelley et al. 2009). Further research must address the informational asymmetry between users and service providers and identify factors influencing comprehension and voluntariness in online consent agreements (Gharib 2022).

In conclusion, obtaining valid consent is critical for digital product development but faces various challenges. By simplifying language, providing granular consent options, offering just-in-

time notices, and ensuring freely given consent, organizations can address the privacy paradox and informational asymmetry, developing more transparent, user-friendly consent processes. This approach benefits both users and service providers. At the same time, further research remains vital to refining solutions that bridge the gap between users' expectations and their actual experiences with consent in digital product development.

Processing Personal Data Based on Legitimate Interests

Processing personal data based on legitimate interests can serve as an alternative to consent when developing digital products, exceptionally when consent may be questionable due to power imbalances or other factors (Article 29 Working Party, Guidelines on Consent, 7) or when obtaining consent is neither practical nor feasible (European Union 2016; Bygrave 2017; Lynskey 2015). Legitimate interests can help mitigate risks associated with obtaining consent, such as revocation or needing separate consent for each specific data processing purpose (Malgieri and Niklas 2020).

However, relying on legitimate interests as a legal basis for processing personal data entails various requirements, such as meticulously balancing the data subject's interests against the controller's legitimate interests (Article 6(1)(f) of the GDPR) and ensuring appropriate safeguards to protect the data subject's rights and freedoms (Article 6(4) of the GDPR) (Niedermeier and Mpame 2019). Additionally, it involves addressing related concerns, including the subjective nature of the balancing test, potential abuse, inadequate transparency, ambiguity, and accountability (Zuiderveen Borgesius 2014; Solove 2013).

Several measures can be taken to address these challenges, such as providing comprehensive guidelines and examples for the balancing test to reduce subjectivity and ensure consistent results (Kamara and De Hert 2018). Enhanced transparency is crucial, with organizations providing concise and easily accessible information about their legitimate interests and the specific purposes for data processing. Strengthened accountability is essential, with organizations adopting robust privacy policies and practices, emphasizing their commitment to respecting data subjects' rights and freedoms (Zuiderveen Borgesius 2014). Privacy by design and default, integrating privacy considerations into system, product, and service design and operation, is crucial (Bygrave 2014). Organizations should support data subjects exercising their rights, such as access, rectification, erasure, and the right to object to processing based on legitimate interests (Kamara and De Hert 2018; Solove 2013).

By adopting these measures, organizations can effectively navigate the challenges of processing personal data based on legitimate interests, ensuring compliance with privacy regulations, and upholding data subjects' rights and freedoms. It is essential to satisfy the specific prerequisites of legitimate interest processing, such as balancing data subjects' interests and demonstrating necessity (European Union 2016). A two-stage approach can harmonize the risks and benefits of using consent and legitimate interest as legal foundations for data processing (Bygrave 2014). By formulating specific and transparent processing purposes without excluding compatible purposes, organizations can address concerns surrounding legitimate interests. Ultimately, these measures contribute to a balanced approach that respects individual privacy rights while facilitating necessary data processing activities in digital product development.

Balancing Consent and Legitimate Interest in Digital Product Development

In digital product development, choosing between consent and legitimate interests as a legal basis for processing personal data necessitates meticulously evaluating factors, including data processing nature and the relationship between the data subject and controller (European Union 2016). GDPR states that legitimate interests are not absolute and mandates controllers to ensure their interests do not override data subjects' rights and freedoms (Kamara and De Hert 2018). GDPR's Recital 47 specifies that legitimate interests can serve as a legal basis if they do not infringe on data subjects' rights and freedoms. Nonetheless, economic interests alone are insufficient for justifying legitimate interests (Esposito 2022); a case-by-case balancing of factors is required.

Legitimate interests might be justifiable if valid reasons exist beyond economic interests and are balanced carefully (European Union 2016; Kamara and De Hert 2018). For instance, if data subjects benefit from improved product safety, it can support justification for the processing. Kosta et al. (2013) argue that legitimate interests should be interpreted considering data subjects' rights and freedoms and the data minimization principle. The balancing test demands that controllers demonstrate legitimacy, necessity, and non-infringement of data subjects' rights. Determining the legal basis for data collection necessitates a thorough assessment of processing circumstances, purposes, and legal framework. Consent should be obtained for specific purposes, but the European Commission established in 1992 that no hierarchy exists among legal bases, and consent is only one alternative.

In digital product development, analyzing the necessity of processing activities and the existence of compelling legitimate interest is essential. This analysis may yield different outcomes,

such as security-relevant products or personalized advertising. Controllers should consider legitimate interests a more ambiguous concept subject to interpretation, making it a less 'safe' alternative for data controllers (Schemer et al. 2014).

In conclusion, consent and legitimate interests are critical in privacy law. A thorough analysis of the legal basis for data processing is essential to ensure compliance and safeguard individuals' privacy rights.

Changing Legal Grounds

Changing the legal grounds for processing personal data requires careful consideration and documentation. To switch the legal basis of processing compliantly, the data controller must ensure the new legal basis is compatible with the original purpose and that any additional legal requirements for the new basis are met (European Data Protection Board 2019; ICO 2020). It involves reviewing the original consent or legitimate interest and assessing compatibility with the new purpose (European Union 2016). If the new legal basis involves consent, the data controller must obtain new, explicit consent from the data subject for the new purpose (Art. 6 GDPR) and inform the data subject of the change in legal basis, allowing them to object (Art. 13-14 GDPR). Suppose the new legal basis involves legitimate interests. In that case, the data controller must conduct a legitimate interest assessment to determine if the new purpose is necessary and proportionate (Recital 47, GDPR), weighing the controller's legitimate interests against the data subject's rights and freedoms (Kamara and De Hert 2018).

Like in the first case, the data controller must be transparent about the change in legal basis and inform the data subject of their rights concerning the new basis (European Data Protection Board 2019). Changing the lawful basis for processing personal data may affect data subjects' rights; therefore, changes must be carefully considered and implemented in compliance with privacy laws and data subjects' rights.

Available Data for Digital Product Development – a Decision Tree

A decision tree can guide the process to ensure legal and ethical compliance when analyzing personal data for digital product development. First, it is necessary to determine whether the data was processed based on consent or legitimate interest. If it was processed on any legal basis other than these, it is generally assumed that it can only be used for developing digital products with

further verification. It is also important to consider privacy requirements such as purpose limitation and evaluate whether they apply to all data or only some. Legal obligations and the validity of the legal basis must be checked if processing methods change, or new purposes arise. Throughout the process, respecting individual privacy rights and expectations remains crucial.

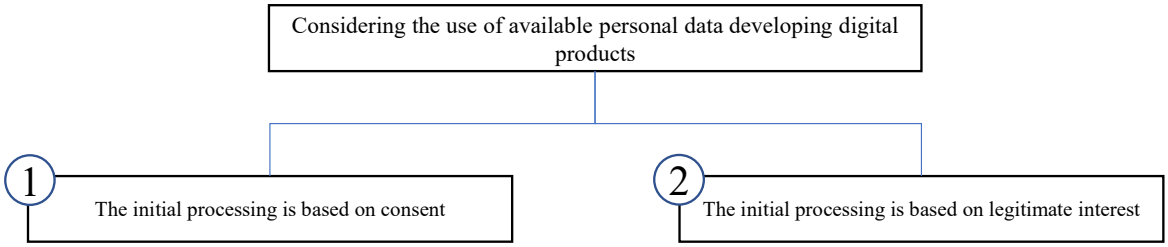

Figure 6: Initial Step: Differentiation by initial legal basis.

If consent is the initial basis for processing, the following steps should be followed when considering the use of data for digital product development.

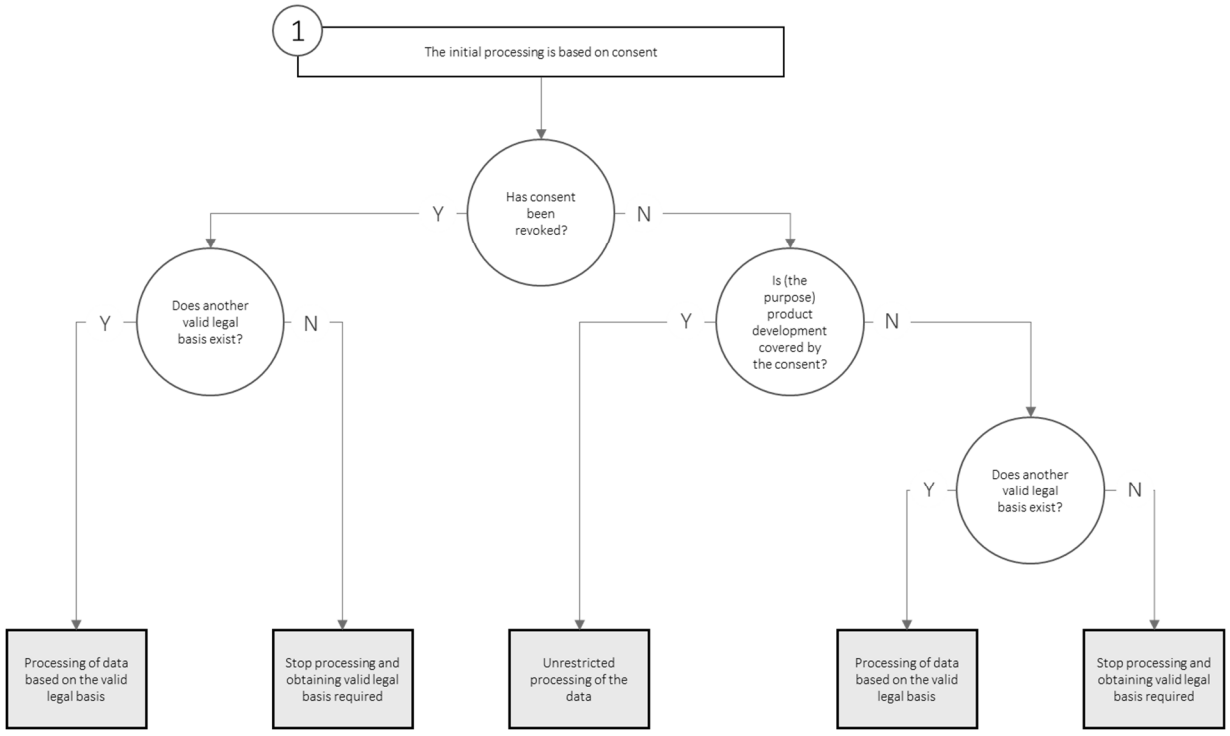

Figure 7: Considerations for Consent-Based Data.

If legitimate interest is the initial basis for processing, the following steps should be followed when considering the use of data for digital product development.

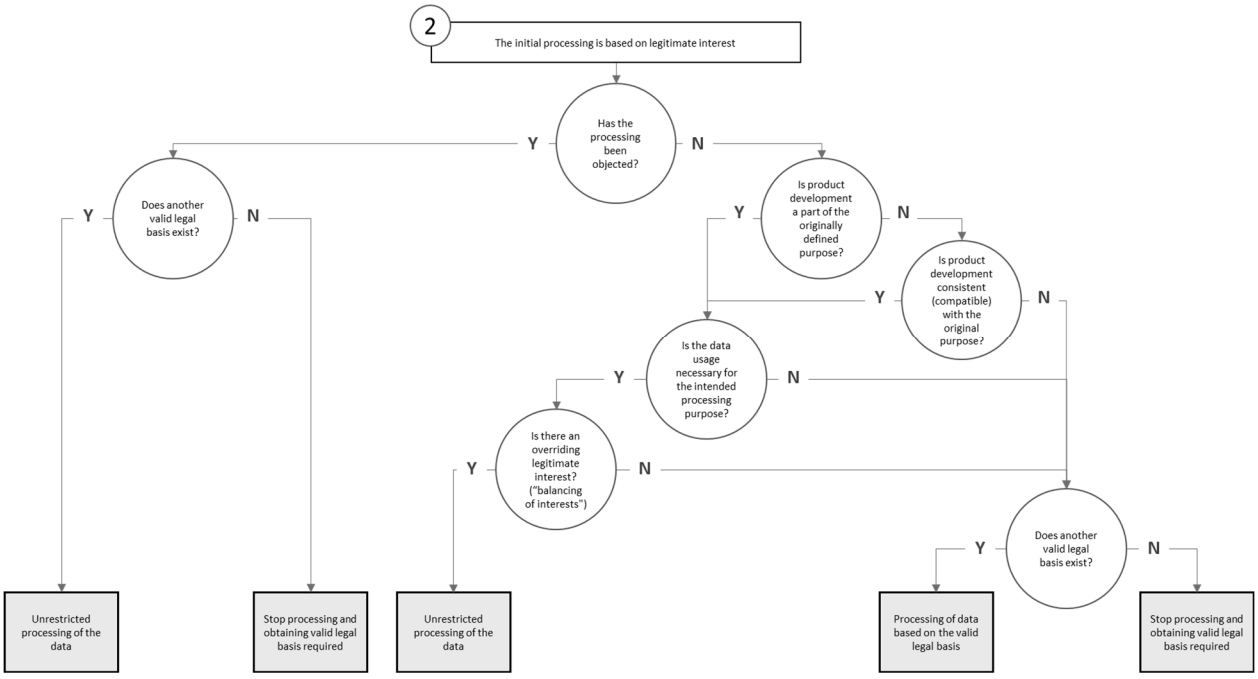

Figure 8: Considerations for Legitimate Interest Based Data.

The outlined steps guide determining whether existing data can be utilized the development of digital products. Acknowledging that each use case requires a unique decision, considering all relevant factors is crucial. Consequently, aspects not depicted or only partially represented in the decision tree may also need to be considered.

Expert Survey on Consent and Legitimate Interest

This section presents the results of an expert survey that aims to gather opinions from privacy professionals regarding the flexibility and permanence of the legal bases consent and legitimate interest from the data controller's perspective. The survey includes 24 IAPP-certified (International Association of Privacy Professionals) privacy professionals conducted during a privacy workshop. The expert survey was employed to enhance the quality of the study, despite the small panel size.

Participants rate their flexibility and permanence responses for two scenarios on a 1-10 scale (1=very low, 10=very high). Flexibility denotes adaptability in processing data without extra action, while permanence pertains that the legal basis remains valid across the processing. Scenario 1 involves consent-based, and Scenario 2 involves legitimate interest-based processing. Additionally, participants evaluate each legal basis's alignment with privacy regulation objectives

("promote transparency, accountability, and security while empowering individual data control") on a 1-10 scale (1=strongly disagree, 10=strongly agree) (see Table 6).

Table 6: Descriptive information (n=24).

	Consent Flexibility	Consent Permanence	Legitimate Interest Flexibility	Legitimate Interest Permanence	Consent meeting privacy objectives	Legitimat Interest meeting privacy objectives
Mean	8,88	6,88	7,21	8,54	8,79	6,67
Med.	9,00	7,00	7,00	9,00	9,00	7,00
Mod.	9,00	7,00	7,00	9,00	9,00	7,00
Var.	0,636	0,897	1,042	0,781	0,694	1,101
Min.	7,00	5,00	6,00	7,00	7,00	5,00
Max.	10,00	8,00	9,00	10,00	10,00	8,00

The non-parametric Wilcoxon Test shows that the scores of “Consent Flexibility” are significantly different from those on “Legitimate Interest” and “Consent Permanence” (see Table 7).

The expert survey indicates that data controllers view consent as more flexible and better suited to meeting privacy regulation objectives. At the same time, legitimate interest offers increased permanence in the data processing

Table 7: Non-parametric Wilcoxon test.

	Consent Flexibility vs. Legitimate Interest meeting privacy objectives	Consent Flexibility vs. Consent Permanence
Test statistics	0,000	0,000
Standard error	32,435	33,491
Standardized test statistics	-4,255	-4,479
Asymp. Sig. (2-sided Test)	0,000	0,000

During panel discussions, we identified several factors as contributing to these results. Data controllers perceive consent as more secure because it allows them to define data use in detail and allows data subjects to provide specific and informed consent for processing. In contrast, legitimate interest involves a balancing process executed by the controller, which might be subject to scrutiny during audits or objections. Furthermore, consent is believed to grant individuals greater control over their data, aligning more closely with privacy regulation objectives.

Results

In this study, we highlight the importance of personal data in digital product development and emphasize responsible data collection and usage. We explore transparency and purpose limitation challenges when incorporating vast data sources and big data technologies in product development. We provide valuable guidance for businesses seeking to balance innovation and privacy concerns in today's digital landscape. We analyze the legal foundations of legitimate interest and consent, identifying key considerations for implementation and potential changes in a legal basis.

To improve transparency and user-friendliness in consent processes, organizations can simplify language, provide granular consent options, offer just-in-time notices, and ensure freely given consent. This approach benefits users and service providers while encouraging further research to bridge the gap between user expectations and experiences. To address challenges in processing personal data based on legitimate interests, organizations can provide balancing test guidelines, enhance transparency, strengthen accountability, incorporate privacy by design and default, and support data subjects' rights. These measures ensure regulatory compliance and balance privacy concerns with data processing activities in digital product development.

When changing the lawful basis for data processing, organizations must carefully consider and comply with privacy laws and data subjects' rights. Our research presents a decision tree to simplify the evaluation of data usage for digital product development, ensuring legal compliance. We gather insights from privacy professionals through an expert survey on GDPR legal bases.

Our survey results show that, compared to legitimate interest, consent-based data processing is perceived as more flexible and aligns more closely with privacy regulation objectives. Consent empowers individuals to exercise greater control over their data. Legitimate interest,

however, requires a balancing process by the controller, which may be scrutinized during audits or objections. Consent is seen as granting individuals more control over their data and aligning more closely with privacy regulation objectives. The average scores for consent and legitimate interest in aligning with privacy objectives are 8.79 and 6.67, respectively.

Limitations & Conclusion

Our study presents several limitations. The results predominantly focus on GDPR, while other jurisdictions receive minimal attention. Likewise, our analysis does not incorporate additional data-related regulations and laws, such as the AI Act and the Digital Services Act. We have not investigated various privacy-preserving or encoding technologies, including pseudonymization, anonymization, differential privacy, and encryption, which can contribute to secure and privacy-compliant data processing depending on the context. The literature employed in our study was not obtained through a systematic literature search, possibly leading to the exclusion of pertinent works. The expert survey conducted for this research included only German participants and expanding the dataset would yield more comprehensive insights. Lastly, examining further questions related to the context and use of legal bases could reveal additional insights and clarify discrepancies between legal regulations and their application.

Future research should address these limitations to enhance the understanding of privacy regulations and best practices in digital product development. Controllers should prioritize transparency and inform data subjects about their data usage, thereby strengthening processing lawfulness, reducing compliance risks, and fostering customer trust. The advancement of digital technologies underscores the need for effective collaboration across various disciplines. Future research should concentrate on refining existing approaches and devising innovative solutions that bridge the gap between legal requirements, technological advancements, and user needs. Encouraging interdisciplinary collaboration paves the way for a harmonious and effective framework that supports digital product development while safeguarding individual privacy rights and fostering public trust in emerging technologies.

5. When Handing out Presents is not Enough! - Influencing Factors on the User´s Willingness to Share Data for Connected Car Services

Published: 25. G-Forum Jahreskonferenz (2020), Innovative Entrepreneurship Session

Authors: Wolfgang Köhler, Christian Schultz, Christoph Rasche

Problem Definition

The automotive industry is facing fundamental technological and economic upheavals. Besides of the technological transition to emission-friendly power units the car becomes a platform, where existing and new companies will compete to offer refined and new services. In this already growing market, data is the fuel for future business models either by digital start-ups or established companies. Especially connected car services (CCS) (e.g. predictive maintenance, entertainment services etc.) promise to generate mayor revenue streams in the future (Seiberth & Gruendinger, 2018). While opportunities to collect vast amounts of data through state-of-the-art technologies are already available, data privacy protection laws aim at giving the user full control of his data at all times. Therefore, the availability of essential high-quality contextual data is mainly limited by the user´s willingness to provide their data voluntarily to companies. Companies who want to gain a competitive advantage need to understand the influencing factors that contribute to the user´s willingness to share data. This knowledge enables companies to take adequate management measures to gain data access. The research guiding question of this study is: What influences the user´s willingness to share data with a company for CCS?

Theoretical Foundation

As digitization fuels servitization and vice versa (Baines & Lightfoot, 2014), car manufacturers have already stopped to imagine themselves as companies that just develop, produce and sell cars. In the medium- to long-term, digital transformation turns the car increasingly into a cyber-physical-system with a multitude of sensors and processing power (Karnouskos & Kerschbaum, 2018). Ultimately, in a time horizon of 8 to 15 years, fully autonomous and connected vehicles have the potential to transform the driving experience by increasing traffic efficiency, reducing pollution, and eliminating up to 90% of traffic accidents (Bonnefon & Shariff, 2016). As the “driver” can spend his time otherwise it is highly likely that the demand for additional services like CCS, especially entertainment offerings, will grow profoundly. A somewhat underestimated challenge every competitor needs to face is the legal compliance in gathering and processing data to offer high-class services. The user´s preferences and influencing factors for sharing data is of high interest for existing or potential service providers to ensure that they have access to high-quality data that fuels their business models. This study fills a gap in the literature on the willingness to share data by users for CCS.

Methodology

The sample stems from an EU-wide survey in August 2017. 5006 persons (2430 male, 2576 female) from Germany, Great Britain, France, Italy and Spain (at least 1.000 persons per country, 18 years or older) participated in an online survey. To answer the research guiding question we test our research model with a multinomial logistic regression. All variables are presented in table 1. The dependent variable consists of 3 categories. The internal construct consistencies for the independent variables, measured by Cronbach's alpha, reach very satisfying levels.

Table 8: Dependent and independent variables.

No	Name	Items	Cronbach's Alpha
A.	**Dependent Variable**		
1	Openness to share data (Categories: Open th share data; Indifferent; Negative on sharing data)	1 (Question 55) 5 point Likert scale	/
B.	**Independent Variables**		
1	Demographic information	gender (male, female) age	/
2	Personal preferences for different areas	5 questions with 5 point Likert scale	,810
3	Perceived personal added value by different services	5 questions with 5 point Likert scale	,846
4	Knowledgeability about the amount of shared data	3 questions with 5 point Likert scale	,838
5	Trust in the data recipient	6 questions with 5 point Likert scale	,927

Results

Regarding the dependent variable 1488 (29,7%) participants qualified themselves as open for data sharing, 1477 (29,5%) as indifferent and 2041 (40,8%) as negative on data sharing. The overall model proves to be significant (2-Log Likelihood: 8987,666; Chi-squared 1868,732, Sig. ,000). The pseudo R-squared measures reach acceptable levels (Cox and Snell: ,312; Nagelkerke: ,352; McFadden: ,172).

Table 9: Classification table.

	Predicted			
Observed	Open to share data	Indifferent	Negative on sharing data	Valid percent
Open to share data	1001	150	337	67,3%
Indifferent	389	248	840	16,8%
Negative on sharing data	308	181	1552	76,0%
Total	33,9%	11,6%	54,5%	56,0%

Overall the model classifies 56% of all cases correctly. Compared to a random classification of cases to the largest individual group (2041/5006=40,77%) the model performs 37,35% (56,0%/40,77%) better.

Table 10: Parameter estimates.

Negative on sharing data [a]		**B**	**Std. Error**	**Wald**	**df**	**Sig.**	**Exp (B)**
Open	Intercept	-1,490	,343	18,846	1	,000	
	Age	-019	,003	45,557	1	,000	,982
	Male	,691	,084	66,969		,000	1,996
	Female	0[b]					
	Personal preferences for different areas	,152	,088	2,961	1	,085	1,164
	Perceived personal added value by different services	,477	,089	28,932	1	,000	1,612
	Knowledgeability about the amount of shared data	-1,312	,063	435,853	1	,000	,269
	Trust	,877	,053	275,862	1	,000	2,404
Indifferent	Intercept	-1,760	,292	36,429	1	,000	
	Age	-,010	,002	16,766	1	,000	,990
	Male	,437	,073	35,782		,000	1,549
	Female	0[b]					
	Personal preferences for different areas	-,031	,075	,177	1	,674	,969
	Perceived personal added value by different services	,387	,074	27,650	1	,000	1,275
	Knowledgeability about the amount of shared data	-,284	,049	33,241	1	,000	,753
	Trust	,433	,044	95,395	1	,000	1,542

[a]The reference category is: Negative on sharing data

[b]Set to zero because of redundancy

Regarding the willingness of users to share data for CCS the main influential factors are knowledgeability about the amount of shared data, trust and perceived personal added value by different areas.

Implications for Entrepreneurship Research and Practice

This study demonstrates that the user´s willingness to share personal data is determined by different factors, like trust, knowledgeability about the amount of shared data and perceived personal added value, that can all be potentially influenced by management measures. It becomes clear that simply handing out a "present" like an app for CCS is not sufficient to ensure the long-term availability of data from users.

Start-ups who want to enter the CCS market might gain a competitive advantage if they follow a nuanced management approach. On the one hand, they need to educate those who are critical of sharing their data and seize measures to build trust. On the other hand, they need to point out the advantages of CCS to those who are already willing to share personal data. Understanding the influencing factors on the user´s willingness to share their data in a fast growing data-driven market will force new and existing companies to attach greater importance to transparency and communication strategies.

This study is based on data from the five largest EU economies. It is possible that in other markets (USA, China), in particular due to different legislation and general attitude to privacy and data protection, the results might turn out differently. Also, other factors, which were not accounted for in this study, may also play a major role regarding the user's willingness to share data. All variables in this study stem from the same questionnaire so common method bias might be an issue. Because of the rather large sample and the clear results we have no indication that the results are distorted.

6. What Determines the Willingness to Share Personal Data? - The Case of the Automotive Industry

Abstract

The automotive industry faces paradigm shifts with the moves to more emission-friendly cars and autonomous driving as the most prominent. Furthermore, the car becomes a platform for services that enables new business models. A prerequisite to make new digital service offerings is the continuous access to the users´ data. Although users might want to benefit from new services, they are not necessarily keen on providing their personal data. Data privacy laws put users in the driver's seat and give them the right to reject data accumulation. To still seize a profit opportunity companies need to secure access to legal user data. The key to amass personal data is that the management addresses the users´ determinants to share personal data voluntarily. The empirical test of our extended privacy calculus model (n=4,440) shows that a considerable share of users perceives greater risk than benefits in data provision. Individual assessment of benefits and risk, data type and trust in the data recipient influences the user´s propensity to disclose information. Targeted management measures to influence the users' sharing affinity are: educating critical users, rolling out trust-building measures and focusing on personal advantages of data sharing.

Keywords: willingness to share data, digitalization, autonomous vehicles, data privacy, user preferences, privacy calculus, data network effect

Submitted for: *Industry and Innovation*

Authors: Wolfgang Köhler, Christian Schultz, Christoph Rasche

Introduction

"The car will become the most complex internet device we have known so far,the car will become a software product."

Herbert Diess (2019, published via LinkedIn), Former Chairman of the Board of Volkswagen Group, second largest car manufacturer at the time in terms of sold vehicles worldwide.

The car has been a setting for human ingenuity for more than 100 years. Carl Benz, an esteemed engineer, patented the first car-like vehicle in 1885. Other inventors like Gottlieb Daimler and Rudolf Diesel developed engines that quickly outclassed the car´s early competitors: the horse and the carriage. Henry Ford paved the way for the modern automotive industry by introducing mass production in the 1910s. A path that led to a worldwide output in 2019 of more than 67 million cars. About twenty years ago, Elon Musk began to disrupt the automotive industry by radically re-thinking the car. Today TESLA is the technological front-runner, competing successfully with luxurious Mercedes cars and widely affordable vehicles by Toyota, Ford, and Volkswagen AG. After a phase of denying the established competitors realized: Although customers still demand individual transportation, the future car will only have minimal commonalities with the car of the 2010s. Performance indicators regarding fossil-fuel powered engines and sophisticated transmissions lose relevance as TESLA models run on electricity with a one gear transmission. However, the alternative, more environmentally friendly powertrain is just the more obvious element of the success story. TESLA nurtures the trend of digitization that turns the car into a cyber-physical-system with a multitude of sensors and processing power (Karnouskos and Kerschbaum, 2017), where data from multiple sources, e.g., devices, platforms, and sensors, is seamlessly processed, exchanged and combined for big data applications. TESLA´s powerful on-board software system can be seamlessly updated through wireless connectivity, just like a smartphone, the internet device of choice for most of us. Consequently, TESLA´s goal is to continuously add new features to its models, with fully autonomous driving as the most prominent future update. In the coming stage of the car´s evolution, the digital features will easily trump the mechanical. The main strategic resource will be data, as it fuels the key digital technologies of the 21st century and powers big data applications and artificial intelligence (AI) (Khatri and Brown, 2010), which enhance existing or ultimately enable new services (Opresnik and Taisch, 2015) and business models (Amir and Zott, 2016; Lamot and Paulussen, 2020; Siddiqa et al., 2016). An

example is connected car services (CCSs) as a relatively new and quickly growing segment with high economic potential, which includes various services to improve the user experience.[1] The key to profit from CCSs is continuous access to personally identifiable information (PII) or "personal data"[2]. This access ensures that CCSs can be provided and improved continuously. Companies formerly not associated with the automotive industry can compete to offer CCS on the car platform. SONY already introduced a prototype at the CES Fair in January 2020, and there are constantly rumors of an APPLE car right around the corner. The continuous sharing of a mass of users is a prerequisite to succeed in the growing market of CCSs. It is an indisputable fact that companies not only have the technological capabilities but a strong incentive to gather as much data as they can to profit from CCSs. But data privacy law puts the user in the driver´s seat for PII and can effectively limit unchecked data accumulation. The growing importance of data and its malicious usage has woken public interest and led to the rapid and continuous improvement of data privacy laws across industries and countries. If a car manufacturer does not want to fall behind, management needs to secure continuing access to legal data by convincing the user base to share its data voluntarily. For this purpose, companies need to expand their focus from organizational and technological processing capabilities (Siddiqa et al., 2016) to managing the users´ expectations, needs and fears dynamically to avoid data sharing consent withdrawals and to motivate users to share their data continuously. Currently, this management task has been neglected or only mentioned as a side note (Gregory, Henfridsson, Kaganer, & Kyriakou, 2020).

The general implications of our results have transfer potential to other sectors and its companies, where PII is a key resource. The stronger the company depends on PII volume and density for its business models, the more critical it is to understand the users´ motivations to share their data in order to keep it accessible.

The guiding research question of this study is as follows: When are users willing to share their data for CCSs? In the first step, we describe the business potential of CCS and the effects of data privacy laws in the automotive industry. Then, we present our extended privacy calculus

[1] According to Seiberth and Gruendinger (2018) CCS include but are not limited to, e.g., personalized settings (seating position, radio) remote services, automated logbook, concierge services, dynamic overview of fuel prices, predictive maintenance, driving style adjustments, online appointment booking, maintenance service, in-car-payment services or different functions-as-a service offerings (additional torque or improved electronic suspension).

[2] The European Commission uses the following definition: "Personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person." European Commission in Article 4 (1) of the GDPR (EU, 2016, 34).

research model and its hypotheses. In the subsequent section, we outline the results. We conclude with a discussion on management implications and research directions.

Theoretical Background

The terms ‘data’ and ‘information’ are oftentimes used synonymously. This false equivalence is the root of fundamental misconceptions of practitioners and academics alike. Data in its purest form are unorganized and unprocessed observations, e.g., raw numbers, figures, images, words, or sounds, derived from observations or measurements. Data are not in short supply in the digital age, where processes are digitized, online behavior leaves traces, and everyday objects are connected to the internet. The International Data Corporation (IDC) forecasts that by 2025, the global datasphere will grow to 163 zettabytes - ten times the 16.1 zettabytes of data generated in 2016 (Reinsel, Gantz, & Rydning., 2017, 3). Unlike most economic goods, data consumption and production are nonrival, and data are not depleted by being used multiple times or simultaneously (Jones and Tonetti, 2018, 1). However, data itself are abstract, with no or limited meaning (El-Amir and Hamdy, 2020) and consequently have very limited economic value. To make data valuable, a company needs to add meaning through contextualization, categorization, calculation, corrections, and condensing (Davenport and Prusak, 2000). Through one or a combination of these activities, data evolve into information, which always has meaning, and its value can even rise through greater use (Tallon and Scannell, 2007). The economic value of data is mainly determined by two factors, the specificity of the data and if the company is allowed to use and process these data through big data analytics (McAfee et al., 2012) to capture value either by expanding or offering new services, to anticipate consumer behavior or to develop new business models (Clough and Wu, 2020). PII is highly specific and possesses value for the development of a variety of digital services. At the same time, its processing does not cause noteworthy additional costs, as there are hardly any significant technological limitations on processing vast amounts of data if the company installs a practical data governance framework (Khatri and Brown, 2010).

While there are frameworks of data governance (Tallon, Ramirez, & Short, 2013) who focus primarily on the technical organizational ability to gather and process data over time, our research focuses on accessing lawfully gathered and processed data to fuel all downstream activities.

CCS as a new business model in the automotive industry

Teece (2018) identifies four imminent major paradigm shifts in the automotive industry: electric vehicles, autonomous vehicles, connected car services[3] and personnel mobility. As the combustion engine loses its attractiveness, car manufacturers will need to discover new unique selling points and sources of income. Ultimately, in a time horizon of 8 to 15 years, fully autonomous and connected vehicles have the potential to transform the driving experience fundamentally by increasing traffic efficiency, reducing pollution, and eliminating up to 90 percent of traffic accidents (Bonnefon, Shariff, and Rahwan, 2016). As the "driver" can spend time not actively driving, it is highly likely that the demand for additional services and entertainment offerings will grow profoundly. The car will become a full cyber-physical-system (Karnouskos et al., 2017, 2) that serves as a service platform with mobility as its core but by far not its single service. On this platform, CCS are a promising market segment (Seiberth et al., 2018). As the portfolio of different CCS expands, the availability, quality and scope of these offerings will likely play an increasingly important role in consumers' choices to buy a certain brand. According to the Strategy and Digital Auto Report (PwC, 2019), the European market, which in 2020 had a CCS market potential of approximately 2.5 billion Euros, will amount to a market size of approximately 14 billion Euros in 2030. The market potential for CCS in Europe, the USA and China is expected to reach a volume of 8.9 billion Euros in 2020 and 73 billion Euros by 2030 - an eightfold increase in only ten years. This would make China the largest market for these services, ahead of the USA and Europe. Providers for CCS are able to collect high-quality real-life PIIs, e.g., geolocation data, vehicle-specific data (acceleration, torque, fuel consumption maintenance messages) and other data that encompass the user's entertainment preferences, web browser histories, synchronized images and videos (Soley, Siegel, Suo, & Sarma, 2018, 11). The potential economic value of this information is already substantial and will most likely grow in the future (Karnouskos et al., 2017). Soley et al. (2018) argue that there is an apparent conflict of interest in how to use the data between principals (users) and companies (agents). While the principals might rather want to limit the amount of PII that the agents receive, the agents have the means and the incentive to gather as much data as they can.

[3] Teece (2018) uses the term "connected cars", which corresponds to connected car services in this study. We therefore continue to use the term connected car services (CCSs).

The effects of data privacy laws

To capture the high potential economic value of PII for CCSs, companies need to meet the second precondition articulated by McAfee et al. (2012): the lawfulness of the data transactions and compliance with the necessary data privacy laws. Overall, data privacy regulations are one area of data regulation laws (see Figure 9), with national security and business protection laws as the others.

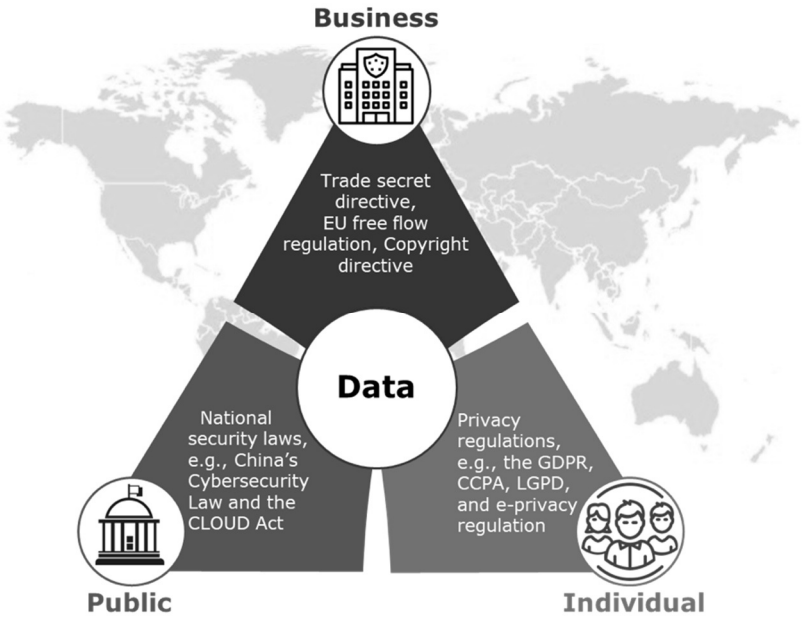

Figure 9: Areas of data regulation (Köhler et al., 2022).

From a macro perspective, Porter and van der Linde (1995) argue that regulation can provide additional incentives for innovation by encouraging the creation of new technologies, products, and markets and discovering overlooked efficiencies. Changing requirements for handling data requires a rethinking of various processes (Koehler et al., 2022). However, especially in the short term, it is likely that innovations cannot completely counterbalance the cost of compliance (Porter and van der Linde, 1995). Martin, Matt, Niebel, and Blind (2019) describe how regulation simultaneously stimulates and constrains innovation in the distinct case of data privacy.

Bélanger and Crossler (2011) trace back the concept of information privacy to Mason (1986), who predicted that privacy would become a major issue with the coming of the then so-called information age. Westin (1967, 7), in an earlier analysis of privacy, defined privacy as …. *"the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others."* While this definition is still valid,

we follow the widely accepted definition of privacy as the individual's ability to control one self's information (Bélanger et al., 2002). Smith, Dinev, and Xu (2011, 996-997) provide a state-of-the-art overview of related concepts, e.g., anonymity transparency, secrecy, and confidentiality.

The right to privacy has been integrated into basic human rights law. However, the design of data privacy legislation worldwide varies considerably depending on the legislator's strategic aims and cultural embeddedness. While there are more than 130 national data privacy laws worldwide (Greenleaf, 2019), the mainstream literature identifies three main data privacy realms: the European Union, the U.S. and China. The superficial view is that the European Union passed particularly tough data protection laws adopted in the member states in 2018. The U.S., with no comprehensive federal law regulation follows a more or less laissez-faire approach that gives their technology companies a competitive edge. In China, comparable data privacy law does not exist. When taking a closer look, the limits of these assertions are no longer clear cut. Although data privacy is not recognized as an individual right in China, data privacy laws have already expanded since the 1980s. Yao-Huai (2020) opines that data privacy protection will continue to have a character that will adhere to traditional Chinese values that favor the state´s collective objectives over individuals' interests. Nothing makes this point more vivid than the current experimentation with the social credit system, where through close surveillance and big data applications, citizens are rewarded for good and punished for deemed bad behavior. Researchers who argue that China has taken the direction of a third way in privacy protection (Pernot-Leplay, 2020) underappreciate the fact that China is ruled by the communist party and separation of the legislative, executive and judiciary powers is not practiced (He, 2012). In China, the government-controlled economy is instrumental in serving multifaceted national interests. Consequently, Chinese data laws favor defined domestic national champions in using data to develop and exploit new technologies. Restricting data access to preserve individual privacy rights is not a major practical issue for the Chinese legislator.

Conversely, in Europe, and in the U.S., especially California, protecting users' privacy follows a different guiding principle. The narrow interpretation of the principle of informational self-determination, where the user must give consent to data processing and has the right to control its personal information unless there are distinct laws in place, gives individuals almost full control of their data. In 2018 European Union countries introduced a comprehensive framework to guarantee fundamental individual rights for personal data protection. Further standards such as the Japanese Act on the Protection of Personal Information (APPI), the Personal Information

Protection and Electronic Documents Act (PIPEDA) in Canada, or the California Consumer Privacy Act (CCPA) are already in force.

From a company's perspective, the unexceptional requirement of lawfulness in processing, storage and management of personal data represents a challenge. The customer´s revocable consent, the performance of a contract, or the exercise of legitimate interest are valid legal bases to process PII. When a user gives consent for certain purposes, the processor cannot exert a subsequent influence on the scope of data processing. The GDPR thus regulates that an extension of processing purposes or collected personal data types requires the renewed consent of the customer (Art. 6 GDPR, Lawfulness of processing and Art. 7 GDPR, Conditions for consent). On the one hand, the individual has the right to easily revoke his consent or object to the processing of personal data at any time. On the other hand, the processing of data beyond its intended purpose requires customers' consent. Furthermore, reprocessing, the transfer of PII to third parties and the application of new processing routines and services are limited. It is safe to say that data protection guidelines and laws regulate the processing of PII more strictly than ever before. Generally, companies that continuously break data privacy laws face harsh direct and indirect consequences. When companies overpay for damages their reputation collapse which leads inevitably to sales decreases and cloudy business prospects. Therefore, companies need to react flexibly to new legal requirements and secure legal access to data simultaneously.

Research Model

In this section, we develop hypotheses to tackle the following question: What determines the user's willingness to share his data for CCSs? At the end of this section, the hypotheses are combined in a research model. To assess this question properly, we draw from different research streams and outline key results in data privacy research, the privacy calculus model, context-dependent user behavior and selected determinants of the widely-cited APCO (antecedents, privacy concerns, outcomes) model by Smith et al. (2011). A body of survey literature confirms the discrepancy between the users´ expressed large concerns about a lack of control over their own data but the rather unrestrictive provision of PII (Rainie and Madden, 2015). For this so-called privacy paradox phenomenon different reasons seem to exist (Baek, 2014; Barnes, 2006; Lanier and Saini, 2008; Smith et al., 2011). The most obvious is that users simply do not care about privacy, especially on the internet (Nissenbaum, 2009). Other explanations are that users are so

accustomed to accepting terms and conditions that they click away the privacy dialogue (Obar and Oeldorf-Hirsch, 2020) or they employ their own data protection strategies, e.g., on Facebook (Young and Quan-Haase, 2013). Hoffmann, Lutz, and Ranzini (2016) outline a deeper psychological explanation for this kind of behavior when they introduce the construct of privacy cynicism. Users have general privacy concerns, but because of a deep feeling of powerlessness against the data gathering of online services, they simply join in. The privacy calculus perspective is a widely accepted model to systematically study the willingness to disclose private information (Dinev and Hart, 2006; Krasnova, Spiekermann, Koroleva, & Hildebrand, 2010; Xu, Teo, Tan, & Agarwal, 2009). Its underlying logic is that consumers perform a risk/benefit analysis of situation-dependent factors to determine if they want to disclose PII (Culnan and Armstrong, 1999). If the perceived benefits outweigh the risks, users provide PII. If this is not the case, individuals handle data release restrictively. This notion has been tested extensively in social networks (Walrave, Utz, Schouten, and Heirman, 2016) and internet contexts, e.g., the decision to install an app on a smartphone (Barth and de Jong, 2017). Dinev et al. (2006) trace the intellectual roots of the privacy calculus framework back to the theory of reasoned action by Ajzen and Fishbein (1980), which later evolved into the enormously popular theory of planned behavior (TPB) (Ajzen, 1991). In general, intentions are the best available predictor of individual intentional conduct, especially if it is infrequent, difficult, or carried out with a significant time delay (Krueger, Reilly, & Carsrud, 2000). The higher the level of preexisting intention is, the more likely it is that the target behavior will occur (Ajzen, 1991). The TPB is a rather general theoretical framework used in different contexts, from sports (Song and Park, 2015) to childhood obesity (Andrews, Silk, & Eneli, 2010) to entrepreneurship. In some contexts, intention is a medium to strong predictor of behavior, e.g., drinking and driving (Armitage, Norman, & Conner, 2002). In other contexts, the research results are mixed, e.g., in entrepreneurship education (Schultz, 2022). The lack of contextual adjustment might explain why the empirical results sometimes do not seem to correspond with the general theoretical reasoning of the TPB (Collins and Carey, 2007). The basic hypotheses that capture the privacy calculus phenomenon are as follows:

Hypothesis (H1). *Individuals will disclose their data if they perceive the personal benefits of sharing information are balanced or greater as the subjective individual risk.*

In their seminal work, Smith et al. (2011, 998) propose an APCO (antecedents, privacy concerns, outcomes) macro model of the relationships between privacy and other constructs. In their literature-based model, trust has a reciprocal effect on privacy concerns and directly affects

behavioral reactions. Castro and Bettencourt (2016) show that trust is a key variable when citizens share data with governmental institutions. Trust is systematically higher for public institutions, e.g., giving up demographic information to a federal agency might not be a problem for most, but granting access to full medical records to a private biotechnology company is more likely to raise concerns. Chellappa and Sin (2005) find that the perceived trustworthiness of the platform or brand positively influences people's willingness to share personal data. The corresponding hypotheses is as follows.

Hypothesis (H2). *Trust influences the willingness to share data for CCSs.*

As a company needs to understand the user´s decision to align its management accordingly, it is also interesting if distinct management measures can actively influence the user´s willingness to share data. Angst and Agarwal (2009), in their study on the adoption of public health records, find that although individuals are concerned about their privacy, their attitude toward sharing can be influenced by message framing. The study uses an experimental design, which incorporates the context. However, its basic result seems highly generalizable. Other studies point in the same direction, e.g., Dinev et al. (2006) indicate that when consumers are informed about a vendor's information practices and perceive the business as fair to them, they are more willing to disclose personal information. The resulting hypotheses is as follows. **Hypothesis (H3).** *Management measures influence the willingness to share data for CCSs.*

The recent study by Winegar and Sunstein (2019) exemplifies the basic challenge of biases in conducting data privacy research. In their survey, the median American consumer was, on the one hand, only willing to pay US $5 per month to maintain privacy. On the other hand, he demanded more than US $80 to give up information. This "super endowment" effect may lead critics to conclude that privacy is not in high regard by the consumer or that American consumers are especially business savvy. However, Winegar and Sunstein (2019) describe different biases that distort the results, e.g., respondents misunderstand the term privacy and do not react consistently regarding the data type they are asked to share. In their seminal early works, Westin (1967) and Laufer and Wolfe (1977), who embedded privacy in a theoretical framework, raise these issues as well. Their main assumption is that the meaning and implications of privacy depend not only on the individual but also on the contextual situation. Therefore, privacy can mean something totally different to one user in a certain situation than to another user in a similar context, which leads to the following hypotheses.

Hypothesis (H4). *The data type influences the willingness to share data.*

Hypothesis (H5). *The data recipient influences the willingness to share data.*

Against the background of information boundary theory, Xu, Dinev, Smith, and Hart (2008) suggest that privacy concerns depend on different individual characteristics (e.g., personality, demographics). Goldfarb and Tucker (2012) find that older people are less motivated to share personal data. Demographic criteria are also a part of the prominent APCO model by Smith et al. (2011). The resulting hypotheses is as follows.

Hypothesis (H6). *Different demographic characteristics affect the willingness to share data.*

We combine the hypotheses in the research model below (Figure 10).

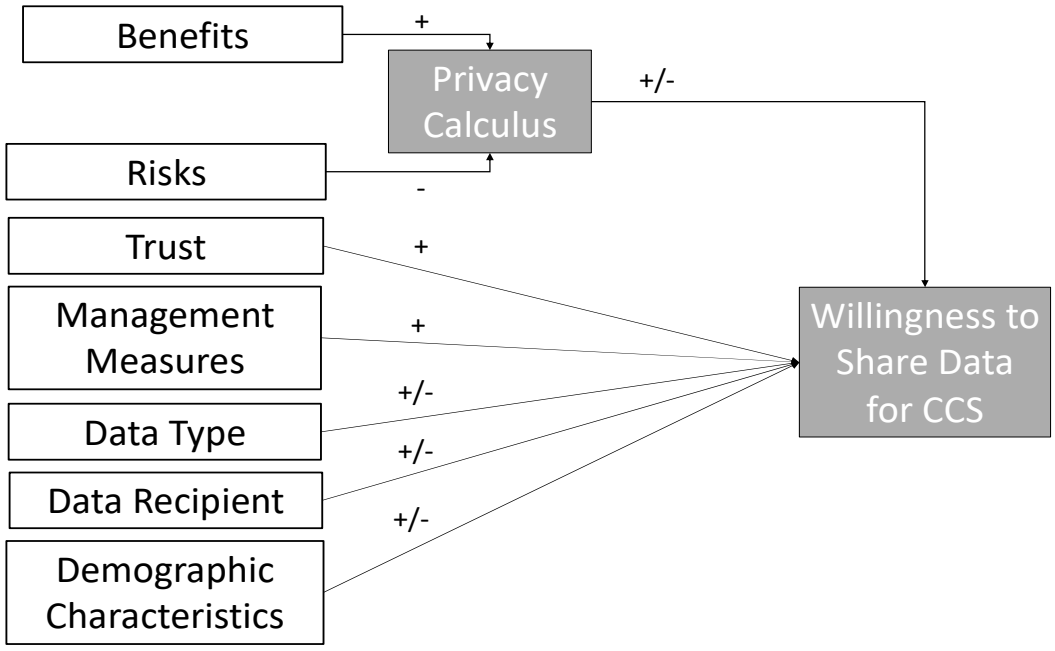

Figure 10: Research model.

EMPIRICAL STUDY

We describe the empirical approach to test our research model (Figure 10) in this section. First, we present the sample. Second, we define all variables and conclude with the results.

Sample

The research agency Research Now on behalf of Deloitte Germany gathered the sample. The survey process started in August 2017 in 5 European countries (Germany, the UK, France, Spain, and Italy) and was completed in October 2017. Slightly more than 1,000 users per country were included in the sample for a total of 5,005 respondents (see Table 12 for demographic

information, additional demographic information is available as supplemental material in a data repository). In this original data sample, 560 persons did not own a car at the time of the survey. As these respondents could only give hypothetical answers on their individual car behavior, they were eliminated to derive a research sample of 4,440 respondents (see Table 11, additional information on the sample is available as supplemental material on a data repository). The respondents were on average 46.61 years old (median 46.00; std. dev. 15.765; max. 91, min. 18).

Table 11: Short descriptive information on the sample(n=4.440).

Variable	Category	n	%
Gender	Male	2221	50.1
	Female	2219	49.9
Monthly net household income	< 1.500€	654	14.7
	1.500-2.499€	1399	31.5
	2.500-3.499€	1217	27.4
	3.500-4.999€	776	17.5
	>= 5.000€	394	8.9
Education level	University degree	1940	43.7
	Qualification for university entrance/A-Level	1448	32.6
	Certificate of Secondary Education	853	19.2
	Primary Education	119	2.7
	No graduation	80	1.8
Country	UK	823	18.5
	France	919	20.7
	Spain	870	19.6
	Italy	944	21.3
	Germany	884	19.9
	<5.000 km/year	626	14.1

Driving kilometers on average per year	5.000-9.999 km/year	1490	33.6
	10.000-20.000 km/year	1799	40.5
	>20.000 km/year	525	11.8
Car Brand (Top 5 of 22 categories)	Renault	471	10.6
	Volkswagen	436	9.8
	Audi	352	7.9
	Asian manufacturers	347	7.8
	Peugeot	341	7.7

Against the background of the critique that the reliance of the bulk of research on information privacy on student-based and USA-centric samples limits their generalizability (Bélanger and Crossler, 2011), our distinct European and comprehensive research sample has the potential to add to the knowledge base on information privacy. There are no obvious biases or distortions in the sample.

3.2 Dependent and independent variables

We follow the general idea of Dinev et al. (2006), who divide the users into three characteristic groups according to their openness to share data (see Table 12). We define the dependent variable as the "willingness to share data for CCSs" (see a similar variable construction by Dimev and Hart (2006) for transactions on the internet).

Table 12: Overview of dependent variable (DV) (n=4.440).

Variable	**Category**	**n**	**%**
Willingness to share data for CCS	Open	1396	31.4
	Indifferent	1330	30.0
	Negative	1714	38.6

Having the privacy calculus model in mind, we construe the dependent variable from question 55 in the questionnaire "Do you want to share your PII and receive an app for your

smartphone?" Respondents answer using a 5-point Likert scale. We define three groups: those who are open to sharing answered with a 4 (agree) or 5 (strongly agree), those who answered with a 3 (indifferent) are assigned to the indifferent group and those who replied with a 2 (disagree) or 1 (strongly disagree) make up the negative group, which is the largest group in the sample. The remaining groups are nearly equally large.

Table 13 provides an overview of all variables, their items and scales.

Table 13: Overview of dependent and independent variables.

No.	Name	Definition	Item(s)
1	Willingness to share data	Willingness to share data for CCS.	1 Item with a 5-point Likert scale ▪ Data sharing to receive free (smartphone, car) apps to access CCS
2	Benefit	Perceived benefit of providing PII to receive CCS services.	5 items with 5-point Likert scales ▪ Benefit for navigation ▪ Benefit for information ▪ Benefit for entertainment ▪ Benefit for safety ▪ Benefit for car management
3	Risk	Perceived risk of opportunistic behavior related to the disclosure of PII for CCS services.	5 items with a nominal (Y/N) scale ▪ Risk of providing navigation data ▪ Risk that the car gets hacked and I lose control ▪ Risk that data is sold to third parties ▪ Risk that my data is sold to an insurance company or a public authority ▪ Risk of unwelcome advertising
4	Trust	Trust that personal information will be handled competently, reliably, and safely.	5 items with 5-point Likert scales ▪ Trust in OEM (Europe) ▪ Trust in OEM (Asia) ▪ Trust in OEM (US) ▪ Trust in IT provider (Europe) ▪ Trust in IT provider (Asia) ▪ Trust in IT provider (US)
5	Management Measures	Management measures that influence the user´s willingness to provide PII for CS.	5 items with a nominal (Y/N) scale ▪ Management guarantees no data sharing with other companies ▪ Management guarantees to uphold the data privacy standard of the industry ▪ Management guarantees transparency of data usage

			▪ Management guarantees no data triangulation with external data sources ▪ Management guarantees data anonymization
6	Demo-graphic Criteria	Age Gender (male/female) Country (France, Germany, Italy, Spain, UK) Car brand, Persons per household, Household income Education level KM per year driving	▪ Age on a metric scale ▪ Gender on a nominal scale (Male/Female) ▪ Country: 1 item on a nominal scale (5 countries total) ▪ Car brand: 1 item on a nominal scale (22 brands total) ▪ Persons per household ▪ Education level ▪ km per year
7	Data Type	Different types of data.	4 items with a nominal (Y/N) scale ▪ Navigation data ▪ Maintenance data ▪ Data of driving behavior ▪ Data on online usage
8	Date Recipient	Different potential recipients of data.	7 items with nominal (Y/N) scale Data transfer to: ▪ Car manufacturer ▪ Car rental company ▪ Insurance company ▪ Authorities ▪ IT provider ▪ Other road users ▪ To other companies in return for payment

Tables 14 and 15 provide detailed descriptive information on all variables.

Table 14: Descriptive information on independent var. measured by a 5-point Likert scale.

Group Affiliation		Open (n=1396)						Indifferent						Negative					
No.	**Variable**	Min.	Max.	M.	Med.	Mod.	Std.	Min.	Max.	M.	Med.	Mod.	Var.	Min.	Max.	M.	Med.	Mod.	Var.
1	Navigation (benefit)	1	5	4,26	4	5	.833	1	5	4.05	4	4	.912	1	5	3.73	4	4	1.173
2	Information (benefit)	1	5	3.93	4	4	.949	1	5	3.48	4	4	1.004	1	5	3.07	3	3	1.240
3	Entertainment (benefit)	1	5	3.64	4	4	1.136	1	5	3.05	3	3	1.102	1	5	2.52	3	3	1.239
4	Safety (benefit)	1	5	4.38	5	5	.838	1	5	4.23	4	5	.903	1	5	3.97	4	5	1.160
5	Car Management (benefit)	1	5	4.24	4	5	.849	1	5	3.97	4	4	.914	1	5	3.70	4	4	1.151
6	Trust in OEM (Europe)	1	5	3.91	4	4	.973	1	5	3.47	3	3	.928	1	5	2.97	3	3	1.126
7	Trust in OEM (Asia)	1	5	3.57	4	4	1.032	1	5	3.12	3	3	.954	1	5	2.70	3	3	1.069
8	Trust in OEM (US)	1	5	3.54	4	4	1.056	1	5	3.06	3	3	.983	1	5	2.63	3	3	1.096
9	Trust in IT provider (Europe)	1	5	3.63	4	4	1.023	1	5	3.15	3	3	.879	1	5	2.72	3	3	1.038
10	Trust in IT provider (Asia)	1	5	3.37	3	3	1.082	1	5	2.88	3	3	.933	1	5	2.51	3	3	1.029
11	Trust in IT provider (US)	1	5	3.58	4	4	1.091	1	5	3	3	3	1.031	1	5	2.58	3	3	1.130

Table 15: Descriptive information on independent variables measured on a nominal scale.

Group Affiliation		Open (n=1396)		Indifferent (n=1330)		Negative (n=1714)	
No.	**Variable**	Yes	No.	Yes	No.	Yes	No.
12	Navigation data (risk)	508	888	365	965	473	1241
13	Car gets hacked and I lose control (risk)	798	598	778	552	794	920
14	Data are sold to third parties (risk)	737	659	871	459	1229	485
15	Data are sold to insurance company/public authority (risk)	405	991	429	901	563	1151
16	Unwelcome advertising (risk)	417	979	513	817	676	1038
17	No data sharing (management measure)	814	582	807	523	1189	525
18	Data privacy standard (management measure)	575	821	426	904	450	1264
19	Transparency of data usage (management measure)	693	703	692	638	758	956
20	No data triangulation (management measure)	427	969	396	934	428	1286
21	Data anonymization (management measure)	476	920	540	790	526	1188
22	External control of guidelines (management measure)	185	1211	220	1110	275	1439
23	Navigation data (data type)	864	532	693	637	554	1160
24	Maintenance data (data type)	974	422	877	453	679	1035
25	Driving behavior (data type)	405	991	429	901	90	1624
26	Data on online usage (data type)	496	900	306	1024	152	1562
27	Car manufacturer (recipient)	960	436	863	467	887	827
28	Car rental company (recipient)	484	912	298	1032	210	1504
29	Insurance company (recipient)	847	549	747	583	1025	689
30	Authorities (recipient)	542	854	418	912	431	1283
31	IT provider (recipient)	496	920	410	920	313	1401
32	Other road users (recipient)	125	1271	59	1271	60	1654
33	To other companies (recipient)	93	1303	60	1270	90	1624

Results

A prerequisite for unbiased results of a multinomial regression is the independence of observations. There is no obvious multicollinearity in the sample, as there are no highly correlated independent variables. No outliers (e.g., age) were eliminated from the analyses, as their exclusion did not have a perceivable influence on the results. Multinomial regression assumes that the dependent variable possesses exhaustive categories. A strong indication that this is the case is that the general model achieves a good overall fit with pseudo-R^2 measurements at .384 (Cox and Snell's R^2) and .433 (Nagelkerke's R^2) and a considerably high loglikelihood at 7549.724. The classification calculations show that the use of this model leads to 60.1percent correct category assignment predictions. Considering the result of a random correct distribution at 38.6 percent (1,714/4,400), this model improves the predictive accuracy by 55.7 percent (60.1%/38.6%). The table below presents the parameter estimates per category (see Table 16). Please note that the negative category serves as a statistical reference for the open and indifferent category, while for the negative- the open category is referenced.

Table 16: Parameter estimates (n=4.440), *p < .01. **p < .05.

(Category 1 (open), 2 (indifferent), 3 (negative), reference category for open and indifferent is the negative category, reference category for negative is the open category, + only significant car brands are reported.)

Cat.	**Variable**		**B**	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**
1	Benefit	Navigation	.026	.646	.137	1	.711	1.026
		Information	.130	.060	4.610	1	.032**	1.139
		Entertainment	.404	.051	63.436	1	.000*	1.498
		Safety	-.059	.069	.726	1	.394	.943
		Vehicle Management	-.090	.070	1.643	1	.200	.914
	Risk	Navigation data is saved	-.070	.111	.394	1	.530	.932
		Car hack/ loss of control	.082	.102	.653	1	.419	1.086
		Data transmission to third parties	.862	.106	66.440	1	.000*	2.367
		Data transmission	.171	.110	2.421	1	.120	1.186

	to public authority						
	Unwelcome advertising	.390	.109	12.918	1	.000*	1.477
Trust	OEM (Europe)	.196	.066	8.665	1	.003*	1.216
	OEM (Asia)	.114	.071	2.571	1	.109	1.120
	OEM (US)	.020	.075	.072	1	.788	1.020
	IT provider (Europe)	.122	.072	2.861	1	.091	1.130
	IT provider (Asia)	.139	.076	3.362	1	.067	1.149
	IT provider (US)	.193	.068	7.998	1	.005*	1.213
Management Measures	No data sharing	.463	.112	17.220	1	.000*	1.589
	Data privacy standard	-.162	.110	2.162	1	.142	.850
	Transparency of data usage	-.163	.102	2.546	1	.111	.850
	No data triangulation	-.224	.112	4.001	1	.045**	.799
	Data anonymization	.056	.109	.263	1	.608	1.057
	External guideline control	-.066	.141	.219	1	.640	.936
Demographic Criteria	Age	-.018	.003	28.446	1	.000*	.982
	Gender	.551	.097	32.105	1	.000*	1.736
	Persons per household	-.007	.023	.099	1	.753	.993
	Household Income	.081	.046	3.098	1	.078	1.084
	Education level	.053	.056	.902	1	.342	1.055
	KM per year driving	.100	.059	2.2820	1	.093	1.105
	UK	.203	.169	1.440	1	.230	1.225
	France	.273	.171	2.567	1	.109	1.315
	Spain	-.389	.166	5.467	1	.019**	.678
	Italy	1.162	.176	43.816	1	.000*	3.196
	Land Rover	1.621	.776	4.365	1	.037**	5.058
Data Type	Navigation data	-.695	.103	45.445	1	.000*	.499

		Maintenance data	-.761	.109	48.560	1	.000*	.467
		Data on driving behavior	-.467	.107	19.166	1	.000*	.627
		Data on online usage	-.741	.126	34.723	1	.000*	.477
		Personal data	-.879	.167	27.801	1	.000*	.415
	Data Recipient	Car manufacturer	-.214	.111	3.726	1	.054	.807
		Car rental company	-.423	.118	12.816	1	.000*	.655
		Insurance company	-.259	.104	6.178	1	.013**	.772
		Government agencies	-.190	.108	3.105	1	.078	.827
		Traffic broadcasters	-.243	.115	4.476	1	.034**	.785
		Other drivers	.074	.217	.116	1	.733	1.077
		Third parties	.546	.213	6.557	1	.010**	1.726
	Intercept		-2.246	.646	12.087	1	.001	
Cat.	**Variable**		**B**	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**
2	Benefit	Navigation	.025	.057	.197	1	.657	1.026
		Information	.020	.051	.157	1	.692	1.020
		Entertainment	.211	.044	23.142	1	.000*	1.235
		Safety	-.032	.058	.303	1	.582	.968
		Vehicle Management	-.100	.059	2.916	1	.088	.905
	Risk	Navigation data is saved	.055	.100	.302	1	.582	1.057
		Car hack/ loss of control	-.012	.089	.018	1	.892	.988
		Data transmission to third parties	.365	.095	14.597	1	.000*	1.440
		Data transmission to public authority	.039	.096	.167	1	.683	1.040
		Unwelcome advertising	.126	.094	1.794	1	.180	1.134
	Trust	OEM (Europe)	.142	.059	56.696	1	.017**	1.152

	OEM (Asia)	.043	.064	.456	1	.499	1.044
	OEM (US)	.068	.068	.987	1	.321	1.070
	IT provider (Europe)	.009	.066	.017	1	.896	1.009
	IT provider (Asia)	.041	.070	.345	1	.557	1.042
	IT provider (US)	.056	.062	.807	1	.369	1.058
Management Measures	No data sharing	.416	.097	18.335	1	.000*	1.515
	Data privacy standard	.016	.099	.028	1	.868	1.017
	Transparency of data usage	-.184	.089	4.295	1	.038**	.832
	No data triangulation	-.226	.099	5.268	1	.022**	.798
	Data anonymization	-.233	.095	6.085	1	.014**	.792
	External guideline control	-.216	.119	3.294	1	.070	.806
Demographic Criteria	Age	-.010	.003	9.974	1	.002*	.990
	Gender	.273	.087	9.925	1	.002*	1.314
	Persons per household	-.017	.026	.446	1	.002*	.990
	Household Income	.004	.040	.009	1	.926	1.004
	Education level	.067	.047	2.025	1	.155	1.070
	KM per year driving	.067	.052	1.670	1	.196	1.070
	UK	.047	.141	.112	1	.738	1.048
	France	.155	.145	1.150	1	.284	1.168
	Spain	-.637	.145	19.188	1	.000*	.529
	Italy	.383	.159	5.758	1	.016**	1.466
	Car brand[+]	/	/	/	/	/	/
Data type	Navigation data	-.464	.091	25.915	1	.000*	.629
	Maintenance data	-.725	.096	57.061	1	.000*	.484
	Data on driving behavior	-.411	.097	17.841	1	.000*	.663
	Data on online usage	-.525	.121	18.810	1	.000*	.592

		Personal data	-.634	.156	16.506	1	.000*	.531
	Data Recipient	Car manufacturer	-.055	.096	.329	1	.566	.946
		Car rental company	-.124	.112	1.228	1	.268	.883
		Insurance company	-.136	.092	2.183	1	.140	.872
		Government agencies	.100	.097	1.061	1	.303	1.106
		Traffic broadcasters	-.219	.102	4.582	1	.032**	.803
		Other drivers	.224	.213	1.110	1	.292	1.251
		Third parties	.568	.198	8.190	1	.004*	1.764
	Intercept		.044	.576	.006	1	.939	
Cat.		**Variable**	**B**	**SE**	**Wald**	**df**	**Sig.**	**Exp(B)**
3	Benefit	Navigation	-.026	.069	.137	1	.711	.975
		Information	-.130	.060	4.610	1	.032**	.878
		Entertainment	-.404	.051	63.436	1	.000*	.668
		Safety	.059	.069	.726	1	.394	1.061
		Vehicle Management	.090	.070	1.643	1	.200	1.094
	Risk	Navigation data is saved	.070	.111	.394	1	.530	1.072
		Car hack/ loss of control	-.082	.102	.653	1	.419	.921
		Data transmission to third parties	-.862	.106	66.400	1	.000*	.422
		Data transmission to public authority	-.171	.110	2.421	1	.120	.843
		Unwelcome advertising	-.390	.109	12.918	1	.000*	.677
	Trust	OEM (Europe)	-.196	.066	8.665	1	.003*	.822
		OEM (Asia)	-.114	.071	2.571	1	.109	.892
		OEM (US)	-.020	.075	.072	1	.788	.980
		IT provider (Europe)	-.122	.072	2.861	1	.091	.885
		IT provider (Asia)	-.139	.076	3.362	1	.067	.870

	IT provider (US)	-.193	.068	7,998	1	.005*	.824
Management Measures	No data sharing	-.463	.112	17.220	1	.000*	.629
	Data privacy standard	.162	.110	2.162	1	.142	1.176
	Transparency of data usage	.163	.102	2.546	1	.111	1.177
	No data triangulation	.224	.112	4.001	1	.045**	1.252
	Data anonymization	-.056	.109	.263	1	.608	.946
	External guideline control	.066	.141	.219	1	.640	1.068
Demographic criteria	Age	.007	.003	28.446	1	.000*	1.019
	Gender	-.551	.097	32.105	1	.000*	.576
	Persons per household	.007	.023	.099	1	.753	1.007
	Household Income	-.081	.046	3.098	1	.078	.922
	Education level	-.053	.056	.902	1	.342	.948
	KM per year driving	-.100	.059	2.820	1	.093	.905
	UK	-.203	.169	1.440	1	.230	.816
	France	-.273	.171	2.567	1	.109	.761
	Spain	.389	.166	5.467	1	.019**	1.475
	Italy	-1.162	.176	43.816	1	.000*	.313
	Land Rover	-1.621	.776	4.365	1	.037*	.198
Data type	Navigation data	.695	.103	45.445	1	.000*	2.005
	Maintenance data	.761	.109	48.560	1	.000*	2.140
	Data on driving behavior	.467	.107	19.166	1	.000*	2.097
	Data on online usage	.741	.126	34.723	1	.000*	2.097
	Personal data	.879	.167	27.801	1	.000*	2.408
Data recipient	Car manufacturer	.214	.111	3.7261	1	.054	1.239
	Car rental company	.423	.118	12.818	1	.000*	1.527
	Insurance company	.259	.104	3.105	1	.078	1.209

	Government agencies	.190	.108	3.105	1	.078	1.209
	Traffic broadcasters	.243	.115	4.476	1	.034**	1.274
	Other drivers	-.074	.217	.116	1	.733	.929
	Third parties	-.546	.213	6.557	1	.010**	.579
Intercept		2.246	.646	12.087	1	.001	

Our results confirm that empirical research endeavors in data privacy need to be context-adjusted. The users' decisions and reactions vary considerably according to the general setting and the distinct questions. Simply put, it just makes a major difference for users if they are asked to share their medical records with their insurance company or their birth date with Facebook. In our case of CCS, context dependency shows up through the significance of nearly all data types and data recipients for the willingness to share data for CCS.

Our results confirm the basic notion of the privacy calculus model that a user calculates the benefits and risks of sharing his data for CCS. Nevertheless, some areas possess statistical significance, while others do not differentiate significantly between the three defined categories (open, indifferent, negative) regarding the willingness to share data for CCS. In the benefits segment, "information" and "entertainment" are relevant. Regarding risk, these areas are "data transmission to third parties" and "unwelcome advertising". As expected, the user's trust is not equally distributed between company categories. OEMs (Europe) and IT providers (US) possess predictive power to categorize users' willingness to share data for CCSs. The management measures of "no data sharing" and "no data triangulation" are solid predictors for user affiliation over all categories. As predicted by the original APCO model, different demographics prove to be relevant as well. As users become older, they are less likely to be open to sharing data for CCSs. Males are more likely to be rather open to sharing data. Surprisingly, some country affiliation plays a role. While Spanish users are rather skeptical, Italian users welcome data sharing. Country affiliation might be a proxy for some underlying cultural determinant. That Land Rover owners are open to data sharing should not be overinterpreted, as only 23 users (0.005%) in the sample drove this brand. The results generally confirm the hypotheses (see Table 17).

Table 17: Overview of the results.

No.	Hypotheses	Confirmed/Rejected
#1	Individuals will disclose their data if they perceive the personal benefits of sharing information are balanced or greater as the subjective individual risk.	**Confirmed** However, not every benefit and risk variable possess a significant influence on the user´s risk/benefit decision for sharing data for CCS.
#2	Trust affects the willingness to share data.	**Confirmed** The user´s trust to selected company types is a significant influence.
#3	Management measures affect the willingness to share data.	**Confirmed** Some distinct management measures possess a significant influence.
#4	The data type influences the user´s willingness to share data.	**Confirmed**
#5	The data´s recipient influences the user's willingness to share data.	**Confirmed** There are some data recipients, who significantly influence the data sharing willingness of the user.
#6	Different demographic characteristics affect the willingness to share data.	**Confirmed** There are some demographic criteria, e.g., age and gender that influence the data sharing attitude. The majority of demographic criteria have no influence.

CONCLUSION

We test an extended data privacy research model on the relationships that determine the users' general attitude to sharing data, which hasn`t been done in the specific context of CCS before. The results demonstrate how multifaceted the user´s decision to provide his PII. A further novelty of our research is that we explicitly consider the role of management measures in influencing users' decisions to share data for CCS. The general implications of our results have transfer potential to other sectors and segments, where personal data is a key resource. The stronger the company depends on personal data volume and density, the more critical it is for the management to understand the user's needs and expectations.

The company's strategic objective is to move users from the indifferent or even negative groups on sharing data to the group that is open to sharing data for CCS. In the logic of the privacy calculus perspective, the management needs to alter the benefit/risk calculation toward a perceived positive balance by emphasizing the benefits and thoughtfully addressing - not downplaying - perceived risk. This can be done either by reframing existing messages (Angst et al., 2009) or experimenting with new messages that target groups might appreciate. From a management perspective, demographic determinants (e.g., gender, age, country) have solely informative value and are useful to channel communication to specific target groups, e.g., in Spain, where users tend to be skeptical about sharing their data for CCS. Conceivable messages of CCS' benefits to users can be potential energy reduction and associated improvement in the climate balance through avoiding congestion or potentially lowering repair costs through efficient maintenance management. Communication reframing the risk areas of "data transmission to third parties" and "unwelcome advertising" might have an especially positive effect. Messages that address these risks and assure users of a high level of transparency in data collection seem viable approaches to decrease the risk perception as well. Regardless, companies should be interested in not sharing data with other companies to develop the best platform, e.g., for autonomous driving, and benefit from the proposed data network effect (Gregory et al., 2020). The proactive execution of distinct management measures can accompany newly designed benefit/risk communication by pointing out that "no data are shared" and "no data triangulation takes place," as our results show that these management measures are of significant relevance to users. An improved communication strategy and the demonstration of adequate actions to guarantee responsible and lawful data handling should increase trust.

Critics might doubt the validity of this study's results as users are asked for intentions, and those intentions – depending on the context – sometimes do not translate into actions (Smith et al., 2011). But in the case of CCS, intentions were the best available proxy to create a large-scale survey, as many of the CCS were not offered at the time, and many others will only be available in the future. Furthermore, the nuanced results show that users have carefully weighted their answers.

A large body of literature is proof of context-dependency in data privacy research. Consequently, it is challenging to derive general conclusions about the underlying mechanisms. Conflicting research results are not necessarily a sign of an inadequate research design, as they show that users react differently to different stimuli depending on the context. Therefore, the generalizability of empirical data privacy results to other contexts will always pose a challenge.

In our case of CCS, the context dependency mentioned above is demonstrated by the significance of nearly all data types and data recipients for the willingness to share CCS data for all user groups. Podsakoff, MacKenzie, Lee, & Podsakoff (2003) describe methods to minimize common method variance bias. We followed their recommendation to vary the circumstances of the surveyed variables and used different response formats of Likert and nominal scales for the dependent and independent variables. The correlation analysis did not reveal any major correlations between variables, which shows at least no obvious indication of a common method variance or other bias in the sample. With 4,440 respondents from 5 different European countries, the sample is rather comprehensive.

Future research endeavors may focus on company-centric aspects of access to legal data as the management process to establish the necessary capabilities. Promising research areas in data privacy include establishing trust over time to a data recipient by the users or the empirical analysis of management interventions that change the willingness of the users to share data. While experiments can bear interesting results, they are not a substitute for the real-life reaction of the users to data privacy issues mainly because of the established context dependency. Especially for understanding change over time inside a company or at the user level, qualitative research, e.g., case studies or action research (Schultz, Mietzner & Hartmann, 2016), are suitable.

7. Data are the Fuel for Digital Entrepreneurship—But what about data privacy?

Abstract

Data have become precious assets in nearly all sectors of the economy. While digital processes, innovative products, and arbitrarily scalable business processes contribute to increased efficiency, higher productivity, and better tailored solutions to customers' needs. Digital entrepreneurs with data-driven business models might even depend on adequate data input for their survival.

The increasing importance of data and the potential for malicious data usage have attracted the public's interest and led to the rapid and continuous improvement of privacy-related laws across industries and countries. Data compliance has become a significant issue for enterprises.

While data privacy is often perceived solely as an obstacle to value-adding processes or as a severe legal risk, the perpetual evolution of data privacy legislation and the need for rule compliance provide business opportunities for entrepreneurs that focus on state-of-the-art digital data privacy solutions.

Keywords: Digital Entrepreneurship, Data Privacy Management, Information Valuation, Information Monetization, Processing Innovation

Published: *The Handbook of Digital Entrepreneurship*, UK: Edward Elgar Publishing Ltd, https://doi.org/10.4337/9781800373631.00027 (Published: 04 Nov 2022)

Authors: Wolfgang Köhler, Christian Schultz, Christoph Rasche

Introduction

The main advancement from digitization lies in collecting, analyzing, and translating formerly unknown amounts of data into relevant information and finally into action as monetization processes (Mikalef et al., 2017). With the rapid introduction of new technologies, digital innovation refers to the continuous adaptation of digital technologies to original market offerings (Nambisan et al., 2017). In a business world that is increasingly digital, new and established companies need to continually evolve their data analytics capabilities and competencies to achieve sustainable business success and perform under platform economics conditions. Due to Porter and van der Linde (1995), regulation can provide additional incentives for innovation by encouraging the creation of new technologies, products, and markets and the discovery of overlooked efficiencies. Changing requirements for handling data require a rethinking of various processes, which opens many business opportunities for digital entrepreneurs, even if "innovation cannot always completely offset the cost of compliance, especially in the short term before learning can reduce the cost of innovation-based solutions" (Porter & van der Linde, 1995).

Data are considered an essential raw material for all key digital technologies of the 21st century. The collection and processing of large amounts of data are crucial for training artificial intelligence (AI) algorithms and advancing the Internet of Things (IoT), where data are seamlessly exchanged between devices, platforms, and sensors (Neely, 2019; Schmidt, 2020) to enable new products and services.

From a digital entrepreneurship point of view, the processing of personally identifiable information (PII) is especially attractive. It enables a variety of digital services and does not cause high additional costs, as there are very few significant technological limitations on processing vast amounts of data. However, not everything that makes business sense and is technologically feasible is also permissible and compliant. From a legal perspective, processing PII is—depending on the jurisdiction—rather highly regulated. Generally, companies that violate data privacy laws face harsh consequences in the form of monetary penalties, a collapsing reputation and a loss of customer trust, which will inevitably negatively affect their sales and business prospects. The widespread ignorance of data protection rules and the resulting fear of being guilty of data misuse poses severe obstacles for organizations in the process of becoming truly data-driven. Management must implement data protection measures that reflect the different legal data privacy spheres (e.g., the EU, the USA, and China) as prerequisites to utilizing value-adding data processes. With different data governance systems,

there is a risk of data opportunism regarding IT-law dumping. Ultimately, effective data privacy management leads to the availability of legal data input for the improvement of digital product quality.

This chapter draws from the relevant literature streams and, especially in the third section, the authors' practical insights from the management of a leading global accounting firm's data risk projects over the last ten years. It provides answers to fundamental questions in data privacy areas for digital entrepreneurs and presents different digital entrepreneurship opportunities. First, we define the term data and differentiate it from related areas, e.g., information or processing, and show what contributes to the value of information and the underlying data. Second, we highlight the significance of data for digital entrepreneurs. Then, we demonstrate how entrepreneurs can cope with data protection laws and use the laws to their advantage.

Data and Information Valuation

The availability and processing of personal data play a critical role in the success of digital platforms and digital service companies and in providing public services ranging from the storage of medical histories (digital patient files) to the creation of digital identity cards. Digital platforms represent new business models that use technology to connect people, organizations, and resources in an interactive ecosystem where exceptional value can be created and exchanged (Parker et al., 2017). Almost every conceivable industry for which information is a crucial component is a prime contender for the platform revolution (Gawer, 2014). Advanced analytics and artificial intelligence can provide consumers with entirely individualized solutions that are delivered in milliseconds (Gawer, 2014). Transaction costs are decreasing due to the increased speed of online dealmaking through digital agents.

What distinguishes data from other resources is its unique characteristics. The production of data through existing sensors or cookies is almost free. Once data are collected, they are not consumed in processing activities as other resources are. Moreover, the value of data is enhanced when insights gained from the data are sustained as a part of an overarching competence-building process, e.g., through the regular exchange of experiences and findings from applied data analytics projects. There is an only minimal variable cost in creating value through processing (Tonetti & Jones, 2020). Unlike physical resources, invisible assets mostly incur no depreciation when employed in value chain activities. The value of data may increase when used in advanced artificial intelligence and data transformation technologies. Through

data combination, machine learning, trained algorithms, pattern recognition, data cleaning and semantic contextualization, data evolve into information, knowledge and competence. Challenging the status of the most valuable companies in the world inevitably leads to a paradigm shift. In the digital era, value creation accrues to brick-and-mortar business models; companies such as Google, Amazon, Alibaba and Facebook take advantage of digital platform economies and scalable invisible asset business models.

On the surface, data are not in short supply in the digital age, where processes are mainly digitized, online behavior leaves traces, and everyday objects are connected to the internet (Internet of Things) (Reinsel et al., 2017). The International Data Corporation (IDC) forecasts that by 2025, the global data sphere will have grown to 163 zettabytes (that is, a trillion gigabytes)—i.e., ten times the 16.1 ZB of data generated in 2016. However, a mass of data does not automatically translate into a huge benefit for a company, for example, in anticipating consumer behavior (Knape et al., 2020).

What are data, information, and processing?

The terms 'data' and 'information' are frequently used synonymously in a wide variety of contexts. The professional and scientific literature supports many different definitions of data and information, depending on the discipline. A fundamental definition of data is that data are unorganized and unprocessed facts, e.g., raw numbers, figures, images, words, or sounds, derived from observations or measurements. Data by themselves do not possess inherent meaning (El-Amir & Hamdy, 2020). The transformation of data to information occurs by adding value in terms of meaning, relevance, and purpose (Davenport T. H. and Prusakv L, 2000). Data by themselves are abstract and can be meaningless, while information always has a meaning. Value is added, according to Davenport and Prusakv L (2000), to data when data are:

- Contextualized *e.g., the purpose for which the data were gathered is known*
- Categorized *e.g., the units of analysis and critical components are known*
- Calculated *e.g., the data have been analyzed*
- Corrected *e.g., errors have been removed*
- Condensed *e.g., the data have been summarized*

Policymakers and legislators worldwide increasingly regulate the processing of data. In this context, it is essential to understand what "data processing" is. According to the European

Commission and based on Articles 4(2) and (6) of the General Data Protection Regulation (GDPR), processing covers a wide range of operations performed on personal data, including through manual and automated means. It includes collecting, recording, organizing, structuring, storing, adapting or altering, retrieving, consulting, using, disclosing by transmission, disseminating or otherwise making available, aligning or combining, restricting, erasing, and destroying personal data. The GDPR applies to personal data processing wholly or partly by automated means and by nonautomated means if part of a structured filing system.

Information and underlying data are some of the most important assets of many companies. Lee et al. (2017) show that it is increasingly important for investors to know how well companies handle data. Data are considered as non-asset under accounting standards and are therefore not recognized as an asset on balance sheets. Data are intangible assets with unique characteristics that preclude the assignment of a universally accepted value. This realization is an important starting point for an overview of all information valuation approaches. Publicly available data are public goods for which no market price is determined. According to the SECI knowledge model (Nonaka et al., 2000), public and explicit knowledge can be transformed into private, implicit, and tacit knowledge through sophisticated learning trajectories. Likewise, Kollmann (2016) observes that the difference between physical and electronic value creation activities lies in the special handling of information.

Just collecting enormous amounts of high-quality data is insufficient to generate a monetary or even competitive advantage. Data are valuable in organizations if they are high-quality, functional, and able to be evaluated by decision-makers ("data intelligence"). Data intelligence is, therefore, crucial to organizations' success. Visconti et al. (2017) describe a useful approach to determine the dynamic data value-adding with five stages: (1) creation and collection, (2) storage, (3) processing, (4) consumption, and (5) monetization (Visconti et al., 2017). In the early phases, namely, data creation, collection, and storage, the data value remains low, while in the data processing and data consumption phases, the data value increases considerably. Data-driven business models can create the maximum value from available data in the final stage of monetization. The model shows significant value generation for companies that have suitable business models and monetization-driven organizational planning (Li et al., 2018).

Further, it is not trivial to track or measure the value of data to different users and for different purposes over time because data consumption is non-rivalrous. The use of data by one person or organization does not deprive other users of the ability to use the data or diminish its

value. These characteristics make it particularly difficult to determine the value of and exclusively monetize specific data products' benefits (Slotin, 2018). As a consequence, different data valuation approaches surface.

Another approach focusing on the quality of data provides that the value of unstructured or erroneous data remains low regardless of the quantity of such data. The criteria of the 3-V model and their extension are useful to evaluate data quality (Chen et al., 2014; Freiknecht & Jonas, 2018):

- **Variety**: In what way are the data available?

Data can be structured or unstructured and are either stored centrally or distributed. In the latter case, linking and further processing becomes considerably more difficult. Typical examples of data that are difficult to process are user-generated content on social media platforms, decentralized collected sensor data, and data collected through cookies.

- **Velocity**: How quickly are the data updated?

Data that are updated at very short intervals will only generate substantial benefits if they are evaluated quickly, preferably in real-time.

- **Volume**: How large is the available mass of data?

There is no objective measure for this criterion. Moreover, in an environment with a continually increasing volume of data, it is difficult to set a meaningful limit, as it can be assumed that data become obsolete relatively quickly.

Although the 3-V model is considered the standard, various authors (Lokhande & Khare, 2015; Shim et al., 2015) extend it by including the characteristics of veracity, variability, and value:

- **Veracity**: Are the data unbiased?

It stands to reason that only correct and unbiased data can provide valid results. However, data can also be incorrect or biased under certain circumstances and still be relevant for data processing. Examples of such circumstances include tweets with misspelled names or incorrectly assigned hashtags.

- **Variability**: To what extent do the data change?

The changeability of data depending on the situation is another criterion.

- **Value**: What is the value of the data?

This category primarily refers to the utility value of the data in terms of the results to be derived, which can often be improved by adequate processing procedures.

Many companies fail to understand the value of their existing data assets and the mechanisms that can increase data value. Based on this section's precise delineation of the terms data, information, and processing and its description of the current methods for capturing the value of information and the data on which they are based, the following section discusses possible means of concretely measuring the value.

How can we measure information value?

To capture and consistently harvest information value over time, organizations must first clarify how to evaluate information as an asset and then develop a comprehensive data strategy to drive value enrichment (Deloitte Touche Tohmatsu Services & Inc, 2020). It is paramount that organizations understand their own data's potential value in the respective context and measure it concretely from a management perspective. Different approaches to information valuation exist. While PwC (2019) limits its valuation methods to the income (net cash flow benefits), market (the traded price in an active market), and cost (the cost of reproduction or replacement) methods, the Global Partnership for Sustainable Development additionally accepts the benefit monetization (an estimate of the monetization of the benefits associated with the data product) and impact-based (an assessment of the economic and sociocausal impact of the data's availability on outcomes) approaches. While these methods have their merits, Duncan Alan and Jones Lydia's (2020) measures provide an overview of information valuation and help adequately capture information value in organizations. In their sophisticated categorization they differentiate between external (direct), internal (indirect), and liability measures.

- External or direct economic measures distinguished into "Market Value" and "Economic Value" of information are used when it is possible to clearly measure the creation of financial goodwill. This approach is beneficial if the data at issue are sold or traded as products or services or contribute directly to the business through profit or loss. These measures are useful for organizations that need to know how information assets should be valued relative to other assets and how to invest in their processing, management, and security. Additionally, the value of information in an enterprise plays an increasingly important role in measuring enterprise value for investors in mergers and acquisitions (Duncan Alan & Jones Lydia, 2020).

- Indirect or internal measures subdivided into “Intrinsic Value”, “Business Value” and “Performance Value” of information are particularly applicable when the focus is on identifying opportunities to improve business operations. The effectiveness or efficiency of existing business processes should be positively influenced without affecting the core business. The quality and potential of an information asset versus its actual benefits can be better evaluated to ultimately improve business strategies. Furthermore, indications of the potential for external or direct economic benefits can be provided (Duncan Alan & Jones Lydia, 2020).
- Liability measures “Cost Value”, “Waste Value” and “Risk Value” of information describe a variety of possible negative financial impacts on an organization. Factors such as system failures resulting in data loss, low data quality, data security breaches, noncompliance with data protection laws and other regulations, or faulty processing are just a few examples of possible causes (Duncan Alan & Jones Lydia, 2020). Several challenges inhibit the transformation from concept to reality, making a calculation of the value and risk associated with data a complicated task. The evaluation of various liabilities and risks and the assessment of their short-, medium-, and long-term impacts can be supported by comprehensive analyses. However, the responsibility for determining potential information monetization and business risk based on a comprehensive liability analysis for all data processing steps is distributed across different business roles (Chief Digital Officer, Data Privacy Officer and others) (Duncan Alan & Jones Lydia, 2020).

Management may apply a collection of these information valuation measures to identify and illustrate the value of information assets. The application of information value measures works best if logical data groupings of related information assets are formed beforehand. Each grouping or class (e.g., customer data or employee data) is treated as a portfolio. It may be necessary to combine several different valuation methods. Which measures an organization can adequately use and under which circumstances depends on its business objectives (Slotin, 2018). Recently, Gregory et al. (2020) raise a new point about the role and value of data and use Tesla as an example to demonstrate their concept of a data network effect. The authors´ main point is that the more of the users´ driving data the company agglomerates, the better their AI for autonomous driving becomes, and consequently, the more valuable the Tesla platform becomes to each user. Gregory et al. (2020) formulate different implications for managers to reap the benefits of this data network effect. First, managers need to make sure that adequate

quantities and quality of data are continuously available through effective data governance. Second, companies need to focus on user-centric design so network effects can result in a perceived superior user experience. Third, managers need to consider responsible privacy principles. If the data network effect is real and what long-term implications it might have is an ongoing discussion (Clough & Wu, 2020).

Data and Digital Entrepreneurs

Digital entrepreneurship is a critical pillar for economic growth and innovation and is recognized by many countries as a very important element of economic development and job creation (Block et al., 2018; Zhao & Collier, 2016). Digital start-ups are organizations that fully rely on digital technologies to create and transfer value and to market, deliver and support digital products or online services (Ahrens et al., 2019; Berger et al., 2019; Zhao & Collier, 2016). Digital entrepreneurs develop innovative digital technologies to create new ventures, business models, digital products, or services or to transform the existing businesses (Ahrens et al., 2019; Audretsch et al., 2017).

Currently, digital entrepreneurs possess a wide range of opportunities to introduce efficient business models that use digital technologies. In addition to using digital technologies to improve coordination, communication, planning, and control, digital entrepreneurs use them internally to enhance (in-)tangible forms of decision-making. This approach affects products and services produced by the organization (software and hardware) directly and helps to scale their venture more quickly (Recker & von Briel, 2019). For this purpose, technologies such as AI (machine learning and deep learning), natural language processing, big data analytics, virtual reality, the IoT, 3D printing, or cloud computing are appropriate (Beck et al., 2017; Rippa & Secundo, 2019; Schulte-Althoff et al., 2021). While business models and their underlying technologies undoubtedly develop rapidly, regulations and legislation governing data handling are catching up with this pace worldwide. Data privacy regimes target the usage of PII, likely to be valuable assets, as the subject of legislative action. Firms should make frequent use of nonmarket strategies to achieve competitive advantages regarding regulations, legislation and legal regimes (Rasche, 2020).

In this section, we outline the relevance of data for digital entrepreneurs from different perspectives. Then, we describe regulatory spheres, risks, and how privacy regulation supports digital entrepreneurs' innovations.

As outlined before, the sheer volume of collected data and data triangulation capabilities has increased dramatically. These rapid technological developments create new challenges for businesses of any size, but especially for new companies, digitization creates opportunities as Porter and Heppelmann (2015) "*believe that the exponential opportunities for innovation presented by smart, connected products, together with the huge expansion of data they create about almost everything, will be a net generator of economic growth*" and that *"there will be more innovation and many more businesses."* The environment for digital entrepreneurs is rather positive as the latest technologies allow companies to use data for unprecedented purposes in their operations. For example, in the automotive industry, the car turns more and more into a cyber-physical-system with a multitude of sensors and processing power (Karnouskos & Kerschbaum, 2018), where existing and new companies can compete with their service offerings. One of these relatively new and quickly growing segments with high economic potential are connected car services (CCS) which include various services around an improved user experience e.g., personalized settings (seating position, radio), remote services, automated logbook, concierge services, dynamic overview of fuel prices, predictive maintenance, driving style adjustments, online appointment booking, maintenance service, in-car-payment services or different functions-as-a service offerings (additional torque or improved electronic suspension) (Seiberth & Gruendinger, 2018). The key to profit from CCS is continuous access to the user´s data and a proper business model. Just as the automotive industry, other industry sectors are undergoing an inevitable transformation process where products' added value continuously shifts from hardware product specifications and quality measures to software quality and solutions.

Digital entrepreneurs must identify potential economic sectors in which their ideas are competitive. In areas where digital transformation is far advanced, digital entrepreneurs may face fierce competition with established players. A well-known example is the social media platform market, where new competitors have to deal with market dominance by Facebook, which is so difficult to cope with that even Google, as a resourceful competitor, was unable to establish their service Google+.

For digital entrepreneurs focused on leveraging their potential or monetization, it is helpful to build in-depth knowledge and broad skills related to information valuation methods. Expert knowledge in recognizing the data´s value makes it possible to demonstrate the value of digital products and services to (potential) customers and clients. This contributes to

competitive differentiation and, in the best case, supports the development of additional sources of revenue or new lines of business.

What is the role of data privacy for digital entrepreneurs?

Among other factors, the repeated and extensive misuse of personal data led to a discussion of and sensitivity to data privacy issues in society. A prominent example is an allegation that misuse of PII assisted in Donald Trump's election. In early 2018, the scandal surrounding Facebook and Cambridge Analytica came to light. The defendants exploited data from more than 50 million users without their consent or even knowledge and used it to target political information and fake news with the aim to manipulate public opinion and, as a consequence, the national election of 2016.

The incentives to collect as much data as possible and to transform it into information as quickly as possible to target customers accurately or offer new services, raise serious questions of legitimacy for public authorities, particularly legislators and ethics committees. Especially the privileged access to data by large digital platforms, e.g. Facebook, Amazon, or Tencent strengthens their market position. So monopolistic market structures and consequently unfair competitive practices are serious possibilities that would hurt the consumer. Thus, digital governance regimes on the macro-, meso- and micro-levels are needed to set codes of conduct regarding compliance issues.

The growing importance of data ensures the dynamic development and continuous improvement of laws across industries and countries. Regulations and laws on data protection and privacy are rather vast. Currently, there are more than 130 national privacy laws, and this number is increasing (Greenleaf, 2019). In many respects, these privacy laws are comparable. However, there are still country- and jurisdiction-specific requirements for the collection, processing, transfer, and storage of data, as well as transparency and documentation obligations.

Matching regulatory requirements, on the one hand, and technical developments in business, on the other hand, is a key challenge. Data governance and compliance require the analysis and implementation of a wide range of business areas and processes. Complexity arises in many respects from the diversity and continuous amendment of data regulations. The focus of digital entrepreneurs is rarely limited to one market and, therefore, they are rarely subject to only a single but multiple national data privacy laws. Additionally, customers are not limited to specific areas, services, or products. People cross borders and legal jurisdictions

with possibly contradictory regulations while using smart devices or connected services with ongoing data processing and transmission. Such use cases may have consequences for data controllers and processors; thus, transparency regarding international laws, the application of those laws and their differences are required. Different regulations focusing on protecting subjects such as individuals, businesses, and market or national security often exist in parallel in each jurisdiction. The superficial view is that the European Union passed particularly tough data protection laws that were adopted in the member states in 2018. The U.S. more or less follows a laissez-faire approach which gives their U.S.-based but worldwide active technology companies a competitive edge. And in China, data privacy doesn't exist.

With a closer look, it becomes clear that the limits of these assertions are not clear cut. It is true that in China, data privacy is not recognized as an individual right, but data privacy laws have expanded since the 80s. Yao-Huai (2020) opines that data privacy protection will continue to have a character that will adhere to Chinese values that favor the state´s collective objectives over individual interests. Nothing makes this point more vivid than the rollout of the social credit system, where through close surveillance and big data applications, citizens are rewarded for good and punished for bad behavior. Authors who nurture a discussion if China has taken the direction of a third way in privacy protection (Pernot-Leplay, 2020) underestimate the fact that China is ruled by the communist party and separation of the powers of legislative, executive and judiciary doesn´t exist in practice (He, 2012). Chinese data laws favor defined domestic national champions in using data to develop and exploit new technologies.

The EU GDPR, the California Consumer Privacy Act (CCPA), and the Brazilian General Data Protection Law (LGPD) are prominent examples of privacy laws and bills worldwide that mainly focus on protecting natural persons' rights and freedoms. Their focus is on individuals and the protection of personal rights, mostly informational self-determination and the right to privacy, and the implementation of security and transparency obligations for the processors of personal data. When raising national or regional requirements to the international level, the issue of cross-border data transfer is of fundamental importance. Single laws can affect different entities, even if the primary focus is on one.

Similar to privacy regulations, some data regulations focus on protecting businesses and markets. For example, the EU Directive on the protection of trade secrets, the EU directive on copyright in the digital single market, and the EU regulation on the free flow of nonpersonal data are laws that focus on protecting and supporting businesses or markets. Furthermore, data

localization obligations, which require the local storage and processing of data and the operation of servers and data centers in the respective countries should be mentioned. One well-known example is the Russian Data Localization Law. When collecting personal data, including through the information and telecommunication network, an operator must document the recording, systematization, accumulation, storage, adjustment (update or alteration), and retrieval of the personal data of citizens of the Russian Federation using databases located in the territory of the Russian Federation (see Article 2, paragraph 1 of Russian Federal Law No. 242-FZ). Additionally, parts of the previously mentioned laws serve economic protectionism and national companies' interests. International data regulations with business as the subject of protection are also diverse and continuously changing.

In addition to individual privacy and business emphases, some data regulations shall ensure national security or the protection of the state and the public interests. On the one hand, data protection concerns the area of national security, such as the protection of important organizations whose undisturbed operations can directly affect national security, e.g., organizations in the energy and telecommunications sectors, and in the international context in terms of restricting and controlling access by foreign states to specific information. On the other hand, data must be continuously monitored to ensure compliance with national legislation and to assist law enforcement. In particular, the Clarifying Lawful Overseas Use of Data Act (CLOUD Act), as well as parts of the Cybersecurity Law of the People's Republic of China ('CSL'), should be emphasized here. The CLOUD Act allows US federal law enforcement agencies to compel US-based technology companies, by warrant or subpoena, to provide data stored on servers upon request, whether the data are stored in the US or on foreign soil. The CSL requires critical information infrastructure operators ('CIIOs') to store personal information and important data generated from China's critical information infrastructures. Regulations from this realm may be particularly contradictory to national privacy laws that seek to protect natural persons' rights and freedoms. The aforementioned contradictions exist, for example, concerning third parties' access to data or the transfer of data to third parties (e.g., various authorities).

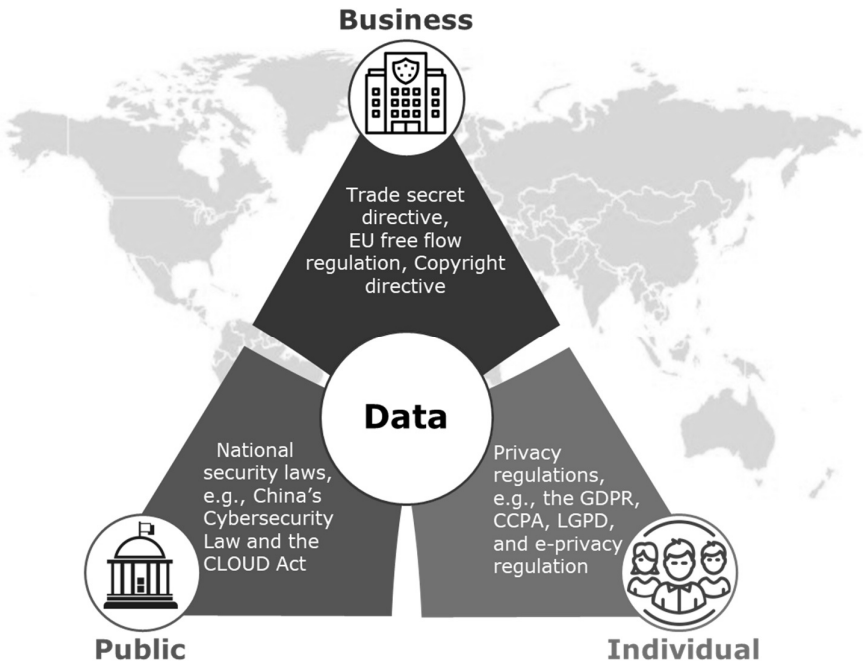

Figure 11: Subjects of Data Regulation.

As outlined above, data can become valuable in different dimensions. It is essential to note that the monetary value of data can only be legally manifested through compliance with data privacy laws, which means a solid legal basis for data processing needs to exist. Furthermore, data can only be shared between different legal entities and further processed at another location (e.g., a head office) under certain conditions.

The narrow interpretation of the principle of informational self-determination (e.g., via consent) and the right of any citizen to control his or her personal information (the right to privacy) place the possibility of realizing the monetary value solely in customers' hands. Transparency and the right to privacy continue to be the focus of legislators' attention in the digital economy. Data protection laws thus also implement ethical restrictions in ongoing macroeconomic processes, the necessity of which is beyond question based on various incidents from the recent past.

Reputation, trust, and, above all, the actual realization of customer control over personal data are decisive factors today and will be in the future. Processing operations that exceed a contract's fulfillment and legitimate interest are, depending on the applicable law, wholly or partly in the customer's hands. Transparency, both internally and externally, regarding data processing practices is indispensable both as proof of compliance and as the operational realization of necessary amendments or actions, e.g., the fulfillment of data subjects' rights.

Noncompliance with data regulations is a significant risk for enterprises in the digitalization era and poses severe obstacles to their transformation into data-driven organizations. The organizational dilemma is that missing out on the potential of data monetization can lead to competitive disadvantages and organizational inefficiencies. In contrast, noncompliance with data regulations can lead to high penalties, market bans, lawsuits, and reputation loss. Corporate political strategies, nonmarket strategies and hybrid strategies can be seen as solutions to this dilemma (Rasche, 2020). The unlawful processing of personal data entails enormous financial risks in the form of claims for damages and fines (Art. 83 GDPR) or even an official ban on the processing of data, threatening the existence of digital products and services. In addition, inadequate protections can increase the risk of a data breach or a deliberate violation, which can have disastrous consequences for a company's reputation. For example, British Airways had to cope with reputational damage after a massive passenger data breach caused by a malicious cyberattack on the BA website between August 21 and September 5, 2018 (ICO, 2020). All of this increases the pressure on those individuals who are responsible for acknowledging data protection requirements, many of which have been in place for a very long time and implementing them in a focused manner. Firms are increasingly challenged to achieve fair, competitive advantages and to engage in corporate social responsibility activities because influential stakeholders and investors are often no longer willing to accept breaches of rules and moral standards.

Martin et al. (2019) describe how privacy regulation simultaneously stimulates and constrains innovation. The researchers identified innovation constraints such as product exclusion in cases of fundamental incompatibility with compliance regulations, entrepreneurial deterrence where concerns related to privacy regulation discourage potential entrepreneurs from starting firms, and as mentioned above, barriers to access to data, especially PII.

Implementing and maintaining regulatory requirements in organizations requires effort and generates costs. Implementing the requirements ties up capital and capacities without directly contributing to value creation, but start-ups may undertake significant innovations in response to regulations—designated compliance innovations. The authors described compliance innovation as changes that make ideas or products compliant (e.g., more privacy-friendly default settings or the use of anonymized PII). The fundamental architecture and value proposition of a product are unaffected, meaning that compliance innovation is primarily about product design. Changes in supply chains (e.g., new partners or suppliers) to ensure that final products consist of compliant components and services are also included (Martin et al., 2019).

Furthermore, the regulation of privacy creates a market for technologies that support data protection and compliance. While regulations impose restrictions on some companies, they also create a potential market opportunity for developing innovative solutions. These solutions can help companies achieve compliance without impacting their regular production and value-added activities. These solutions are sold to others affected by regulations or are developed by companies to cope with their internal compliance activities. Martin et al. (2019) refer to advances in this sector as regulation-exploiting innovations.

Opportunities in Data Privacy for Digital Entrepreneurship

In addition to innovative approaches that can contribute to evaluating, structuring, processing, or analyzing data, a broad business opportunity arises from data processing regulations. Rapidly changing markets, the development of new technologies, and the swift parallel development of regulatory frameworks and legislation are challenges for all market players. With every new requirement, the question arises of who can provide the best and most efficient solution. Current research shows that regulation, in this case, privacy legislation, is not just an obstacle or barrier to start-up innovation.

In this section, we outline opportunities for digital entrepreneurs to benefit from data privacy regulation. We show the importance and advantages of technical compliance innovations before highlighting digital entrepreneurs' opportunities concerning technology, software, and IT innovations.

The GDPR increases the importance of IT security and refers to having 'appropriate technical and organizational measures' (TOMs) in place. Article 32(1) of the regulation states as follows:

"Taking into account the state of the art, the costs of implementation and the nature, scope, context, and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk."

Different security measures are mentioned (see table 18).

Table 18: Security measures supporting data privacy, complied from Art. 32 GDPR.

Security Measure	Examples
Pseudonymization	Replacement of PII by one or more artificial identifiers or pseudonyms
Encryption	Encoding a file so that it can only be read by certain people
Confidentiality	Protecting data against unauthorized or unlawful processing
Integrity	Protecting data from unauthorized changes to ensure that it is reliable and correct
Recoverability	Generation of regular backups and use of data recovery centers
Evaluation	Regular review of TOMs on effectiveness and plausibility

The implementation of these measures is essentially about matching the risks that arise from processing. It is necessary to regularly check whether the used measures are appropriate.

With the trend towards processing increasing amounts of data and more categories of data, and with more processing purposes and techniques, maintaining adequate protections requires the ongoing development of security measures. Simultaneously, new standards and norms for processing data, the development and usage of algorithms, and the protection of processed data continually evolve. State-of-the-art measures are moving targets that require the continuous development of standards within organizations. For entrepreneurs, developers, and existing service providers, these underlying conditions create a permanent need for new products and services.

The GDPR does not prescribe any standards for how to implement the respective protections in an organization. Digital entrepreneurs can develop and implement genuine technical solutions that fulfill these requirements. The areas with increasing demands on data processing and simultaneously increasing processing volumes that can be aided by technical solutions or support tools are incredibly diverse. The following table illustrates selected topics.

Table 19: Advantages of using technical compliance innovations.

Specific advantages of a technological deployment and implementation	Closer view
Category: Transparency	
Optimized identification and structuring of (personal) data as well as complete, effective, and efficient implementation of documentation requirements	By visualizing data streams and processing activities and the associated increased transparency, it is possible to understand data processing within the organization, whether their collection and documentation correspond to current requirements, and whether, in addition, processing activities exist that are not considered and therefore unknown to the organization.
Visualization of data streams and processing activities	
Improved integration with upstream systems and collaboration and exchange between available IT systems	
Category: Efficiency and Productivity	
Increased efficiency and productivity	Efficiency and productivity can be increased by reducing manual workloads and process complexity and by reporting and preventing data compliance incidents. This is achieved due to, e.g., the fact that errors caused by humans, such as accidental deletion or overwriting of data, are prevented to the extent possible and are largely eliminated.
Centralized and standardized process management and data acquisition from various IT systems	
Increase of product quality and process flexibility	
Optimization of decision-making processes	
Cost reduction through simplification/ automation	
Category: Compliance and Liability	
Minimization of compliance risks and follow-up costs	The transparency of the entirety of the existing data processing activities and the associated completeness ensure the ability to check compliance and to make flexible
Faster verification and monitoring of compliance conformity	

Strengthening relationships with customers and business partners as well as the company's reputation	adjustments to the processing activities in the event of any infringements of the applicable requirements. This is achieved by central control in the sense of the uniform collection of all processing operations through complete identification and standardized structuring of all data collected and processed within the whole organization.
Real-time adaptations to changing conditions (legal or internal)	

In addition to the implementation of measures to comply with legal requirements, another perspective is the design and implementation of appropriate processes and systems that support necessary follow-up measures in the event of a data breach. The focus is on immediate measures regarding the identification, assessment, documentation, and reporting of incidents on time and in an appropriate form and scope.

Artificial intelligence can be used to quickly and accurately identify compliance risks and derive early warning information in the event of a data leak or unlawful data processing. Furthermore, Romeike (2019) elaborates on the benefits of machine learning and multilayer learning for maintaining data compliance. The former enables the identification of potentially criminal acts (e.g., money laundering), whereas the latter supports the identification of and defense against previously unknown malware or cyber-attacks (Romeike, 2019). IBM Security and Ponemon Institute (2019) describe the benefits of investing in data breach detection technologies in their annual Data Breach Report. The faster data breaches are detected and contained, the lower the resulting costs. Organizations can improve their ability to contain security breach damages through security automation and intelligent orchestration capabilities that provide visibility to security operations centers.

Currently, however, technology, software, and IT innovations focus on the methodology of data processing, which is increasingly being used to support various compliance measures within organizations. In their article, Schadt et al. (2019) address the potential of using “regulatory technology innovations” as artificial intelligence to ensure regulatory compliance with the help of intelligent compliance technologies. They refer to the intersection of requirements and technologies as regulatory technology. This includes compliance solutions that provide efficient added value based on smart technologies and

remediation. Technologies of this type are based on data mining systems, robotic process automation, and predictive analytics and help to efficiently implement the flood of new compliance regulations, monitor regulatory changes, ensure high compliance quality, optimize compliance, and automate compliance testing processes (Schadt et al., 2019).

Wicke and Püster (2019) investigate the opportunities for and risks of using big data, data analytics, and artificial intelligence to meet data protection requirements. The study shows that efficient data use and processing using self-learned algorithms can also be effective with unstructured data. In this way, the balance between regulatory compliance requirements and the technologies' strategic usability can be maintained. Furthermore, the authors noted that data processing regulatory requirements are subject to continuous change and that flexible designs of operational compliance processes using AI algorithms are therefore necessary (Wicke & Püster, 2019). While restrictive data governance regimes and provisions may contribute to lowering cybercrime, they may suffocate entrepreneurial creativity.

In addition to applications that support single areas such as data structuring, data detection, or risk analysis, the market for so-called privacy management tool software to achieve and maintain compliance with privacy laws and regulations is growing strongly. These tools are often modular in design and offered as platform solutions, enabling providers to continually expand their service offerings or adapt them to changing conditions and allowing organizations to purchase technical support tailored to their needs. The modules focus on specific processes, specifications, or requirements of individual laws, such as supporting or conducting privacy impact assessments (PIAs) under the GDPR or focusing comprehensively on the requirements of countries or regions such as the GDPR in the EU or the CCPA in California. In addition, provisions from other laws and guidelines, such as website tracking regulations under the EU Cookie Directive, are integrated into the platforms. The variation of software functionalities is broad and continually evolving. In addition, legal requirements such as documentation and transparency obligations result in an enormous amount of manual effort both during implementation and in maintaining compliance with privacy laws and regulations. Businesses use data privacy management software to automate manual processes, support transparency requirements, and leverage applications to manage their internal privacy programs, centralize control and visualize various organizational compliance processes via dashboard functions.

A few years ago, most of the current privacy management tools did not exist. The increased regulation of data processing and data transmission (especially personal data) creates new unsolved problems.

Table 20: Development of privacy management software providers between 2018 and 2019 (turnover development from O'Leary, 2020).

Company	**Year established**	**Revenue** [million USD]	**Turnover development** [2018-2019]
OneTrust	2016	283.2	+ 141.6 %
TrustArc	1997	83.7	+ 20.7 %
BigID	2015	42.0	+ 92.6 %
Securiti.ai	2018	39.7	+ 3,506.4 %
Crownpeak	2001	24.3	- 23.2 %
WireWheel	2016	12.5	+ 108.2 %
Exterro	2004	12.4	+ 44.5 %

A need for new solutions can arise directly from changes in legal or regulatory frameworks, so start-ups with data-driven business models or data processing and management capabilities should closely follow legislative developments. A current example of a regulatory change resulting in a need for action is the so-called "Schrems II" ruling in which the European Court of Justice declared the "Privacy Shield," a formerly suitable guarantee for sending EU citizens' data to the USA and further processing it there, invalid. Those who implemented the regulation were able to follow how the legislators work on new standards because of the ruling, with a considerable impact on many businesses.

In general, there are many opportunities for digital entrepreneurs to develop technical solutions for businesses to meet the growing challenges of data processing, structuring, valuation, and monetization. New big data technologies, such as AI, process mining, data mining, and predictive analytics are increasingly used in new applications to ensure data compliance and to develop new approaches and solutions. Digital entrepreneurs can satisfy new requirements through known and possibly adapted methods or through the development of new techniques and technologies. It is safe to say that the market for data privacy will

continue to grow since individuals and institutions want to preserve their digital gestalt and identities.

8. Mapping the Field Across Disciplines: Data Protection Research in Law, Economics and IT

Abstract

Digitalization is recognized as one of the most important societal changes of the last decades and influences more and more aspects of our personal lives. While digital products and services are easily accessible, the accumulation of personally identifiable data and its triangulation by AI has potential negative effects, such as discrimination against social groups or manipulation of consumer behavior or public opinion. As a result, protecting personally identifiable information (PII) has become an essential issue for various stakeholders, such as users, companies, and lawmakers, who sometimes pursue conflicting goals. In addition to legislation, jurisprudence, and corporate privacy activities, many academic publications on data regulation have emerged. The development and implementation of the General Data Protection Regulation (GDPR) in the European Union is a major milestone in regulating PII collection and an important reference point in the academic discussion of the last two decades. To understand the scope, depth, and nuances of the academic discourse on data regulation research, we need to consider the relevant disciplines. We use a simple conceptual framework to map the knowledge base across the law, IT, and economics disciplines. The results indicate that research on data regulation is growing. However, each discipline has unique growth patterns. Notably, the most cited papers are predominantly from the IT domain, highlighting its central role in the academic discourse on data regulation. The study also reveals a significant shift in research topics across stages and disciplines, demonstrating data regulation research's dynamic and adaptive nature. We identify a clear temporal output lag between disciplines, reflecting their different responses to the regulatory shift. This finding provides valuable insights for theorists and practitioners alike.

Published: Not yet published. Working paper in preparation for submission.

Authors: Wolfgang Köhler

Introduction

In our rapidly evolving digital landscape, data-related legislation is emerging as a critical interdisciplinary concern with far-reaching implications for numerous academic disciplines and research areas. Building on decades of remarkable progress, privacy research has predominantly flourished within discrete academic fields (Bräunlich et al., 2020). However, in light of these advances, we must emphasize privacy research's inherent interdisciplinary nature. The widely cited study by Smith, Dinev, & Xu (2011) pointed in this direction. However, the emergence of new data regulations, particularly the General Data Protection Regulation (GDPR), which came into force in 2018, has transformed several disciplines in academic and practical contexts. It is important to recognize that the nature of the far-reaching influence of this legislation stems from its multidimensional nature.

On the one hand, data legislation is fundamentally a legal issue characterized by legal requirements, legal interpretation, data compliance, and legal advice in business. At the same time, it intersects with the field of Computer Science and Information Technology (henceforth referred to as 'IT'), where the focus is on predominantly technical issues related to data handling, processing, categorization, transfer, deletion, systems, and tools (Lenhard, Fritsch & Herold, 2017; Akil, Islami, Fischer-Hübner, Martucci & Zuccato, 2020). Furthermore, it is deeply intertwined with business management, focusing on business processes, risk, compliance management, and privacy procedures. Examining law, IT and economics reveals only isolated instances of interdisciplinary initiatives and research. Despite the apparent interdisciplinarity, the literature shows only intermittent intersections among these disciplines. Given this broad-spectrum impact, a set of intriguing questions arises:

1. How are new laws, such as the GDPR, perceived by rather loosely connected academic disciplines such as law, business administration & economics, and IT?
2. Do academic disciplines exhibit a time-lagged increase in research interest?
3. If a time delay exists, can it be logically justified, or does it indicate a lack of discipline-specific research in individual disciplines?

The answers to these questions not only influence our understanding of the current academic landscape on data regulation but also have significant implications for future data-related issues (Shu & Liu, 2021), such as the future relevant laws and acts from the EU Data Strategy. They also have practical relevance for organizations seeking efficient compliance management, moving from understanding to implementing requirements, systemic support (efficiency), and proactive risk and compliance management. In addition, the insights from this research could

inform organizational strategies, such as assigning data-related laws and their requirements to specific areas of expertise or responsibility within the organization (e.g., assigning GDPR responsibilities to the legal department). It could also inform the case for involving other relevant areas in the early stages of compliance and policymaking.

In this paper, we find answers to these questions through a bibliometric study that maps the research field across the disciplines (Zhang, Wang, de Pablos, Tang & Yan, 2015) of law, IT, and economics. We shed light on the interplay between data-related legislation, various academic disciplines, and business practices. We seek to provide valuable insights that could guide future legislation (Gao, Wu & Yang, 2022), research directions, and business strategies in this critical area of data legislation.

Theoretical Background

The birthplace of innovation is often the intersection of different research disciplines, and the collaborative interplay, mutual enrichment, and collective evolution of these fields are highly encouraged in both academia and practice. When examining data and privacy-related legislation such as the GDPR, a striking observation is the apparent paucity of interdisciplinary studies in privacy research; a surprising finding given the numerous scholarly works highlighting the interdisciplinary nature, importance, and impact of privacy regulation and other data-related legislation (e.g., Smith, Dinev & Xu, 2011; Kitsiou, Tzortzaki, Kalloniatis & Gritzalis, 2021; Labadie & Legner, 2023).

Definitions

The contemporary understanding of privacy, particularly as it relates to information privacy, is inextricably linked to the development of information technology. While the concept of privacy spans multiple disciplines, it is argued that the research contributions of information systems (IS) scholars have significantly influenced the current conceptualization of information privacy. This influence is expected to continue as information privacy continues to evolve (Smith, Dinev & Xu, 2011).

The relationship between the evolution of privacy and technological advances is indelible, a connection that has been thoroughly explored in a wealth of scholarly literature (Gasser, 2016). These scholarly works highlight how privacy as a normative concept has continuously adapted to the introduction of new information and communication technologies (Bennett, 1992; Regan, 1995; Smith, 2000; Solove & Schwartz, 2020; Vincent, 2016; Westin,

1967, as cited in Gasser, 2016). This shift has been evident since the early modern period, a time when urbanization and the spread of mass communication began to reshape the conventional landscape of face-to-face interactions (Gasser, 2016).

Privacy laws and regulations, such as the General Data Protection Regulation (GDPR), have significant implications, especially for businesses and various organizations. In particular, they require companies and organizations to initiate or modify their business processes, rethink their data collection, processing, transfer, and deletion procedures, and create and deploy effective solutions. Such requirements inherently underscore the intersection between privacy laws and business considerations, making the connection to the field of economics apparent.

Our study elucidates two critical phases in the progression of GDPR: the preparation phase and the implementation phase. The preparation phase starting with a notable judgement by the European Court of Human Rights (ECHR) in I v. Finland, no. 20511/03 on July 17, 2008. This landmark decision emphasized the paramount importance of practical security measures in safeguarding personal data, and ignited the legislative progression towards an innovative regulation, thus reshaping and significantly modernizing the data protection landscape. The preparation phase concluded with the broad agreement on the final text of the GDPR in 2015.

The implementation phase started with the formal publication of GDPR in 2016, followed by the enforcement of its provisions in 2018. This phase encompasses relevant legal precedents and continued enforcement efforts up to the year 2022. This phase characterizes the ongoing process of integrating GDPR into the practical sphere, marking the transition from theoretical legal constructs to tangible implementation. To further illustrate the interrelationships between GDPR-related events, we summarize GDPR legislative milestones (see Table 21).

Table 21: Development steps of the GDPR.

Year	Milestones	Stage/ Phase
2008	• European Court of Human Rights: Failure to take effective information security measures to protect sensitive personal data violates right to privacy – I v. Finland, no. 20511/03, 17 July 2008	**A** Preparation
2011	• July 22: Opinion by the EDPS on data protection in the European Union.	
2012	• January 25: Proposal for strengthening the digital economy and rights to privacy on the Internet. • March 7: The EDPS adopts an opinion on the Commission's data protection reform package. • March 12: Opinion on the proposed data protection reform. • October 5: The Article 29 Working Party brings further aspects into the debates on data protection reform.	
2014	• March 12: The European Parliament adopts the GDPR	
2015	• July 15: A general approach to the GDPR was achieved. • July 27: EDPS recommendations on the final text of the GDPR. • December 15: Agreement on the GDPR.	
2016	• February 2: Publication of an action plan for the implementation of the GDPR. • May 24: The Regulation enters into force.	**B** Introduction/ Enforcement
2017	• January 10: EU Commission proposes two new regulations (ePrivacy and Regulation 45/2001 on EU institutions).	
2018	• May 25: Application of the General Data Protection Regulation.	
2018 - 2022	• Case law and enforcement	

Conceptual Framework: Size, Time, Place, and Composition

This paper aims to map the research on privacy regulation across disciplines. As scientific discourse is an important part of understanding the mechanisms in the development of an issue (White, 2004), it is not the focus of this study. A promising way to study discourse thoroughly is Critical Discourse Analysis (CDA) by Fairclough (1995), which proposes a three-dimensional framework with text, discourse, and social practice as its cornerstones. Since we aim not to analyze discourse, we have chosen the more appropriate analytical framework by Hallinger and Kovačević (2019). In their influential paper on scientific mapping of educational administration research, the authors propose a conceptual framework to systematize the published literature in their target field. This very basic idea of structuring has been adopted, either in its entirety (Hallinger, Gümüş & Bellibaş, 2020). or with minor additions, in other bibliometric studies in other research fields (Tekdal, 2021). This framework is easy to understand, provides a comprehensive picture of the knowledge base, and has already proven

itself in various fields. The framework consists of four different dimensions: size, time, place, and composition. As measured by the number of published research papers, size encompasses empirical and conceptual research on the topic and is relevant to capture the state of academic development, e.g., growth in publications indicates research interest. Although measuring size does not provide specific insight into quality, a critical mass of publications is required before results can be synthesized into usable knowledge (Pilkington & Meredith, 2009). "Time" in this review refers to publication trajectories over specific periods. To monitor long-term changes in the development of the knowledge base. Large differences in growth over time could indicate a shift in scholarly engagement and the development of particular topics (Gumus, Bellibas, Esen & Gumus, 2018). "Space" is defined as the geographic distribution of texts in the literature. Understanding how scholarship is distributed globally can reveal academic networks and successful research collaborations (Oplatka & Arar, 2017). Composition often refers to the "intellectual structure" of the knowledge base. Elements of intellectual structure include disciplinary composition, major research topics, and patterns of interrelationships (Zupic & Čater, 2015).

Hypotheses

There are widely cited bibliometric papers that do not formulate explicit hypotheses (e.g., Glänzel & Schoepflin, 1999; Thanuskodi, 2010) and proceed in an exploratory manner to report findings and draw conclusions about the state of the field from the results. This approach is perfectly legitimate and may be appropriate for some fields. However, the lack of falsifiable hypotheses may compromise the perceived level of academic rigor. Furthermore, the generated theoretical implications may be far less understandable and transparent to readers in retrospect.

Since we can draw on the results of existing bibliometric studies in the field of data management and regulation, we opt to formulate hypotheses as preliminary falsifiable assumptions that will be tested through our research (Akil et al., 2020; Ducato, 2020). As bibliometrics is widely regarded as a quantitative research method, formulating hypotheses seems appropriate (Vijayakumaran, Rahim, Ahmi, Rahman, & Mazlan, 2020).

The first set of underlying general hypotheses is that the volume of publications in data regulation research has grown (Hypothesis A), and this growth has not been evenly distributed (Hypotheses B, C, D, E). Furthermore, the scientific discourse through publications on specific topics develops differently depending on the discipline and stage. The publication activity across disciplines resembles a wave structure, where the publication output first grows in the discipline of law and then triggers publications in the disciplines of IT and economics. The

second set of hypotheses concerns the geographical location of publications (hypotheses F) and the composition in terms of topics (hypotheses G). Table 22 provides an overview of the hypotheses and the corresponding indicators.

Table 22: Hypotheses and Indicators.

No.	Concept	Hypotheses	Indicator
A	**Size**	A. The publication output on data regulation has grown strongly.	Overall publications development
1			
2		B. The publication volume develops differently per discipline.	Publication per discipline (Law, IT, Economics) per year
3		C. The growth in publication output is different per discipline.	Smoothed publication growth per discipline per year
4		D. Most cited papers stem from the Law discipline as they are important references for authors in other disciplines.	Overall most cited papers 1-20
B	**Time**	E. Different publication waves are recognizable. Beginning in the law discipline, a wave of publications picked up and triggered a growing wave of publications in the disciplines of IT and economics.	Publications per discipline and stage
5			
6			Overall distribution of publications per discipline per year
7			Distribution per discipline and stage
8			Smoothed publication growth per discipline per year
9			Most cited paper per discipline and stage
C	**Space**	F. The main collaborations take place between the EU and the US. Chinese data privacy researchers don't collaborate that often with researchers from other cultural backgrounds.	Overall geographical collaboration overall
10			
11			Geographical collaboration per stage
D	**Composition**	G. Different topics of data regulation research are dominant depending on the discipline and stage.	General text clouds
12			Text clouds per stage
13			Text cloud per discipline and stage

Bibliometric Analysis

The author follows the guidelines of Aguinis, Ramani & Alabduljader (2018) to provide a fully transparent picture of the research process and its underlying decisions. To assess the hypotheses (Table 22), we undertake a bibliometric analysis of GDPR-related scientific publications. Bibliometrics is the most suited methodology to map this topic across disciplines as it visualizes (Mou, Cui, & Kurcz, 2019) and quantifies the relevance of authors and

documents by statistical methods and can also reveal thematic structures (Ellegaard & Wallin 2015; Thompson & Walker, 2015; Koo, 2017). Zupic & Čater (2015) provide a state-of-the-art overview of bibliometric methods in the management discipline.

Sample Description: Scientific disciplines, databases, and stages of publications

In this paper, we systematically evaluate the contributions from distinct research disciplines toward the understanding of GDPR. Web of Science was chosen as the preferred database among various options due to its expansive access to a range of databases, including Web of Science Core Collection, Inspec, KCI—Korean Journal Database, MEDLINE, Preprint Citation Index, and SciELO Citation Index. In the subsequent analyses in this study, we draw from two data samples. The first, termed the 'Large Database,' includes all selectable databases in the Web of Science. It incorporates a total of 3,485 entries from 1999 to 2023, with 3,315 entries specifically falling within our analyzed timespan from 2008 to 2022. The second, the 'Small Database,' constitutes a smaller, more concentrated dataset, encompassing 389 entries from 2012 to 2023 and precisely 366 entries within our analyzed period from 2012 to 2022. In the opening section of our paper, we employ the Large Database to analyze the broad scope of publication numbers. However, due to certain missing parameters in this database, it could not be efficiently analyzed using Bibliometrix, a tool specifically designed for in-depth bibliographic analysis (Aria & Cuccurullo, 2017). As we progress into the subsequent sections, we thus transition our analytical focus to the Small Database. This strategic shift serves three essential purposes:

Improved Discipline Segregation: The Small Database offers a more separation data, which bolsters the accuracy of our disciplinary analysis.

Advanced Bibliographic Analysis: The Small Database, with its complete parameters, facilitates the use of Bibliometrix for a comprehensive bibliographic analysis.

Robust Hypothesis Testing: The opportunity for detailed bibliographic examination provided by the Small Database enables us to test our hypotheses with increased rigor, fostering the extraction of more nuanced insights from the data.

In essence, this strategic transition to the Small Database augments the precision and depth of our investigation, enabling a more refined exploration of GDPR-related research across the three disciplines.

In this paper, we divide our analysis into two distinct phases, as demonstrated by the accompanying charts and tables. For the disciplines of IT and economics, our analysis solely comprises phase B, as there is no publication data available for these disciplines during phase A. In contrast, the discipline of law exhibits data across both phases, indicative of an earlier engagement with GDPR-related topics within this field. This phased approach provides a structured lens to evaluate the evolution of GDPR-related research across different disciplines.

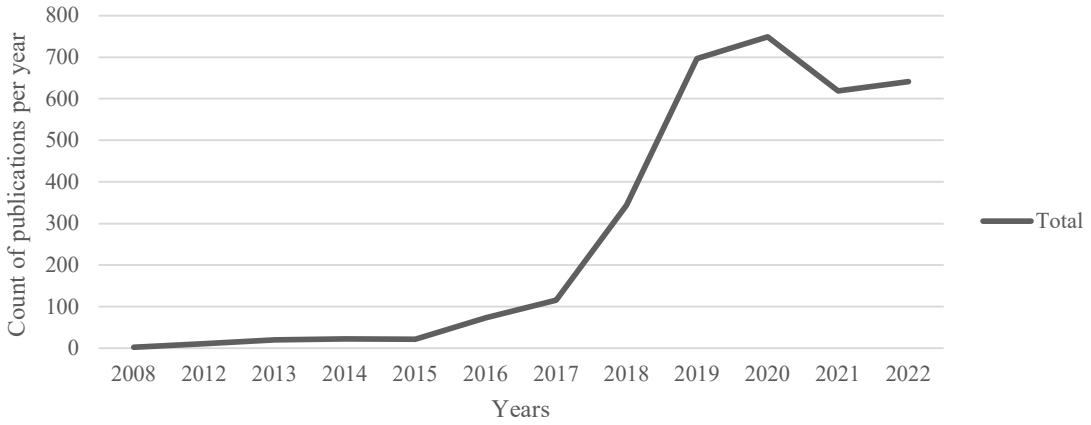

Figure 12: Overall publication development per year in total (large database; n=3,315).

Figure 12 presents the temporal progression of annual publication rates from 2008 to 2022, with the X-axis denoting the timeline from the earliest recorded point in 2008 through 2022 and the Y-axis quantifying the number of publications per annum.

The graph illustrates a nuanced narrative of publication trends. In the initial phase until 2015, we observe a subtle but steady growth in publication counts, exclusively driven by the discipline of law. However, from 2016, and more noticeably from 2017 onward, the publication trajectory took a sharp turn upward, suggesting an intensified research momentum, indicating critical developments in the field that spurred scholarly interest resulted in a boost in publication numbers. The year 2021 marks a deviation from this trend, with the graph showing a noticeable dip in the number of publications. However, the following year, the publication rate began its upward climb once again without regaining its previously accelerated pace.

Bibliometric Findings

Figure 13 provides an insightful examination of the annual publication rates within the different research disciplines of economics, law, and IT, specifically in relation to GDPR-related studies. The data shows that IT is the discipline with the highest total publication count over the years, demonstrating an intense focus on GDPR within the field. Conversely, economics has the

lowest aggregate publication count, suggesting a more restrained scholarly activity. Chronologically, law was the pioneering discipline, with GDPR-related publications appearing as early as 2008. IT followed suit in 2012, and economics entered the GDPR literature landscape in 2013.

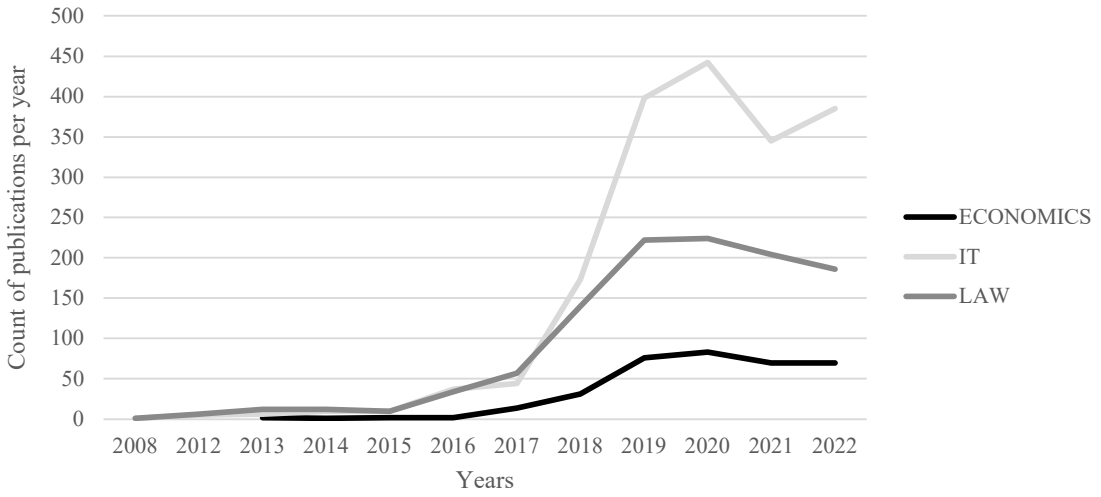

Figure 13: Publication per discipline (law, IT, economics) per year (n=3,315).

An interesting pattern emerges when looking at growth rates across disciplines. Until 2017, IT and law had comparable trajectories of publication growth. After 2017, however, IT's publication numbers accelerated dramatically, marking a deviation from law's growth rate. As for the discipline of economics, a noticeable increase in the number of publications did not appear until 2016, although it lagged significantly behind the counts of law and IT. The year 2018 stands out as a key moment for all disciplines, with an apparent peak in GDPR-related publications. This suggests that 2018 may have been a landmark year for GDPR-related research, which warrants further investigation to understand the cause of this spike. The observed trends raise intriguing questions about each discipline's differential engagement with the GDPR and underscore the need for interdisciplinary dialogue and collaboration to effectively address this sweeping regulation.

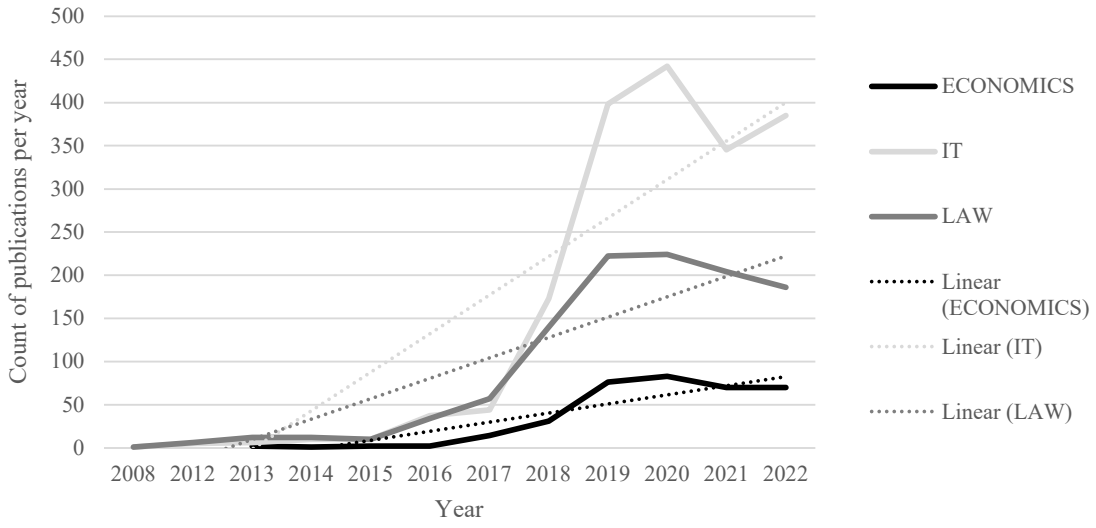

Figure 14: Smoothed publication growth per discipline (law, IT, economics) per year (n=3,315).

Figure 14 shows the annual smoothed growth rates of publications in the fields of law, IT, and economics. The graph shows a clear growth trend in the number of publications for all disciplines. It is noteworthy that IT shows the steepest growth curve, suggesting a rapid expansion of GDPR-related research in this discipline. On the other hand, while also showing growth, economics has the most modest rate of increase, indicating a relatively slower pace of research output. These trends highlight the relative research intensity of the different disciplines and underscore the growing importance of GDPR in these fields, particularly in the IT sector.

Table 23: Overall most cited papers 1-20 between 2008 and 2022, sorted by global citations (n=366).

No	Authors	Title	Year	Journal	Stage/ Phase	Discipline	Local Citations	Global Citations
1	GOODMAN B; FLAXMAN S	EUROPEAN UNION REGULATIONS ON ALGORITHMIC DECISION MAKING AND A "RIGHT TO EXPLANATION"	2017	AI MAGAZINE	B	IT	16	661
2	WACHTER S;MITTELSTADT B;FLORIDI L	WHY A RIGHT TO EXPLANATION OF AUTOMATED DECISION-MAKING DOES NOT EXIST IN THE GENERAL DATA PROTECTION REGULATION	2017	INTERNATIONAL DATA PRIVACY LAW	B	LAW	33	354
3	SELBST AD; BAROCAS S	THE INTUITIVE APPEAL OF EXPLAINABLE MACHINES	2018	FORDHAM LAW REVIEW	B	LAW	6	157
4	TIKKINEN-PIRI C;ROHUNEN A;MARKKULA J	EU GENERAL DATA PROTECTION REGULATION: CHANGES AND IMPLICATIONS FOR PERSONAL DATA COLLECTING COMPANIES	2018	COMPUTER LAW \& SECURITY REVIEW	B	LAW	13	124
5	GODDARD M	THE EU GENERAL DATA PROTECTION REGULATION (GDPR): EUROPEAN REGULATION THAT HAS A GLOBAL IMPACT	2017	INTERNATIONAL JOURNAL OF MARKET RESEARCH	B	ECONOMICS	6	123
6	BERNAL BERNABE J;LUIS CANOVAS J;HERNANDEZ-RAMOS JL;TORRES MORENO R;SKARMETA A	PRIVACY-PRESERVING SOLUTIONS FOR BLOCKCHAIN: REVIEW AND CHALLENGES	2019	IEEE ACCESS	B	IT	0	112
7	WACHTER S	NORMATIVE CHALLENGES OF IDENTIFICATION IN THE INTERNET OF THINGS: PRIVACY, PROFILING, DISCRIMINATION, AND THE GDPR	2018	COMPUTER LAW \& SECURITY REVIEW	B	LAW	5	87
8	DE HERT P;PAPAKONSTANTINOU V;MALGIERI G;BESLAY L;SANCHEZ I	THE RIGHT TO DATA PORTABILITY IN THE GDPR: TOWARDS USER-CENTRIC INTEROPERABILITY OF DIGITAL SERVICES	2018	COMPUTER LAW \& SECURITY REVIEW	B	LAW	17	86
9	KAMINSKI ME	BINARY GOVERNANCE: LESSONS FROM THE GDPR'S APPROACH TO ALGORITHMIC ACCOUNTABILITY	2017	SOUTHERN CALIFORNIA LAW REVIEW	B	LAW	18	84
10	DE HERT P;PAPAKONSTANTINOU V	THE NEW GENERAL DATA PROTECTION REGULATION: STILL A SOUND SYSTEM FOR THE PROTECTION OF INDIVIDUALS?	2016	COMPUTER LAW \& SECURITY REVIEW	B	LAW	9	73

No.	Authors	Title	Year	Journal	Stage/ Phase	Discipline	Local Citations	Global Citations
11	MANTELEYO A	THE EU PROPOSAL FOR A GENERAL DATA PROTECTION REGULATION AND THE ROOTS OF THE `RIGHT TO BE FORGOTTEN'	2013	COMPUTER LAW \& SECURITY REVIEW	A	LAW	6	68
12	MANTELERO A	AI AND BIG DATA: A BLUEPRINT FOR A HUMAN RIGHTS, SOCIAL AND ETHICAL IMPACT ASSESSMENT	2018	COMPUTER LAW \& SECURITY REVIEW	B	LAW	6	67
13	SELBST AD;BAROCAS S	THE INTUITIVE APPEAL OF EXPLAINABLE MACHINES	2020	FORDHAM LAW REVIEW	B	LAW	3	60
14	SCHWARTZ PM;PEIFER KN	TRANSATLANTIC DATA PRIVACY LAW	2017	GEORGETOWN LAW JOURNAL	B	LAW	0	59
15	HAMILTON RH;SODEMAN WA	THE QUESTIONS WE ASK: OPPORTUNITIES AND CHALLENGES FOR USING BIG DATA ANALYTICS TO STRATEGICALLY MANAGE HUMAN CAPITAL RESOURCES	2020	BUSINESS HORIZONS	B	ECONOMICS	0	54
16	KREUTER F;HAAS GC;KEUSCH F;BAEHR S;TRAPPMANN M	COLLECTING SURVEY AND SMARTPHONE SENSOR DATA WITH AN APP: OPPORTUNITIES AND CHALLENGES AROUND PRIVACY AND INFORMED CONSENT	2020	SOCIAL SCIENCE COMPUTER REVIEW	B	IT	0	54
17	KUPERBERG M	BLOCKCHAIN-BASED IDENTITY MANAGEMENT: A SURVEY FROM THE ENTERPRISE AND ECOSYSTEM PERSPECTIVE	2020	IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT	B	ECONOMICS	0	52
18	MOURBY M;MACKEY E;ELLIOT M;GOWANS H;WALLACE SE;BELL J;SMITH H;AIDINLIS S;KAYE J	ARE `PSEUDONYMISED' DATA ALWAYS PERSONAL DATA? IMPLICATIONS OF THE GDPR FOR ADMINISTRATIVE DATA RESEARCH IN THE UK	2018	COMPUTER LAW \& SECURITY REVIEW	B	LAW	11	51
19	KAMINSKI ME	BINARY GOVERNANCE: LESSONS FROM THE GDPR'S APPROACH TO ALGORITHMIC ACCOUNTABILITY	2019	SOUTHERN CALIFORNIA LAW REVIEW	B	LAW	8	51
20	VICTOR JM	THE EU GENERAL DATA PROTECTION REGULATION: TOWARD A PROPERTY REGIME FOR PROTECTING DATA PRIVACY	2013	YALE LAW JOURNAL	A	LAW	3	50

Table 23 provides an overview of the top 20 most cited papers in the disciplines of law, IT, and economics from 2008 to 2022, ranked by the number of global citations. The law discipline dominates this list, with most of the top 20 entries coming from this discipline. However, despite having fewer entries, the IT discipline has the highest citation growth rate and the most cited paper worldwide. Law follows with the second-highest citation growth rate and the second-most cited paper. Economics has the fewest citations of the three.

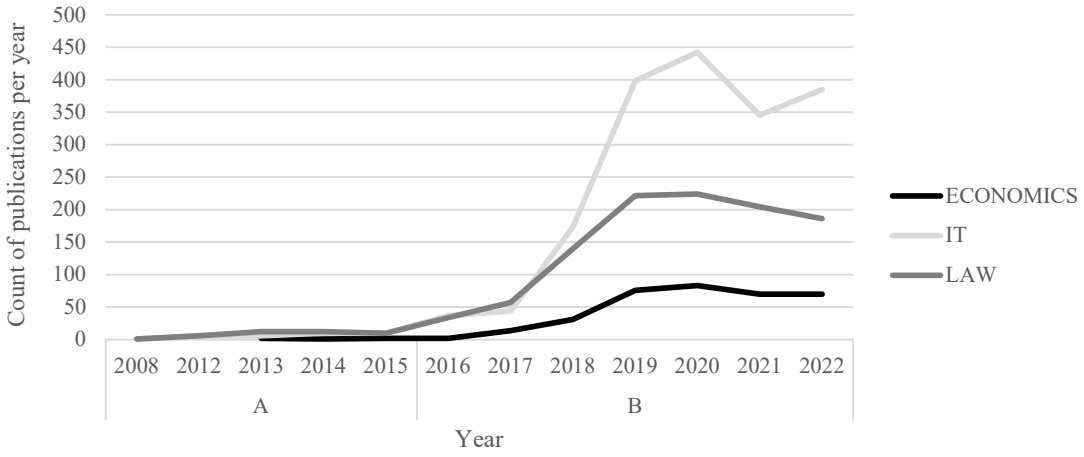

Figure 15: Publications per discipline and stage (n=3,315).

Figure 15 provides a comparison of the publication rates in two different phases, divided by the year 2015. Stage A covers the period up to the end of 2015, while stage B covers the period from 2016 to 2022. The x-axis represents these two stages, and the y-axis represents the number of publications per year for each discipline. During stage A, publication rates across all disciplines appear relatively flat, indicating minimal growth. The transition to stage B marks a significant change, beginning with a noticeable publication increase in 2016. A strong contrast in the number of publications per year can be observed between stages A and B, highlighting a significant intensification of scholarly activity related to GDPR from 2016 onwards. The division into stages A and B allows for a clearer understanding of the temporal trends in GDPR-related research output, signaling the growing importance of GDPR as a research topic after 2015.

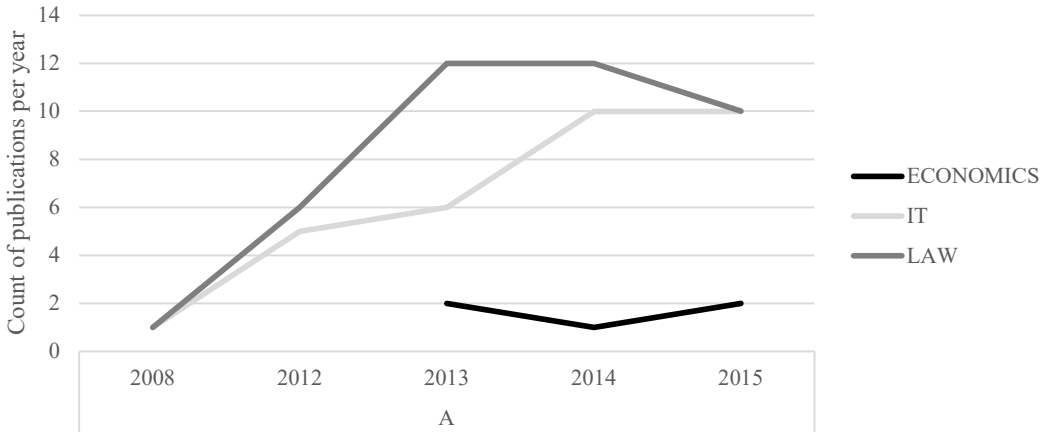

Figure 16: Publications per discipline stage A (n=78).

Figure 16 shows the publication rates within the disciplines of law, IT, and economics during stage A, which covers the period up to the end of 2015. During this period, the law discipline shows the highest growth rate in GDPR-related publications, suggesting a leading role in this area of research. The IT discipline follows closely behind, indicating a growing interest in GDPR research in this field. Conversely, economics research shows a significantly sparse output during stage A, with minimal data available. That may indicate a delayed entry, or less initial engagement, by the economics discipline in GDPR-related studies during this stage. These findings provide valuable context for understanding the early development of GDPR research in these disciplines.

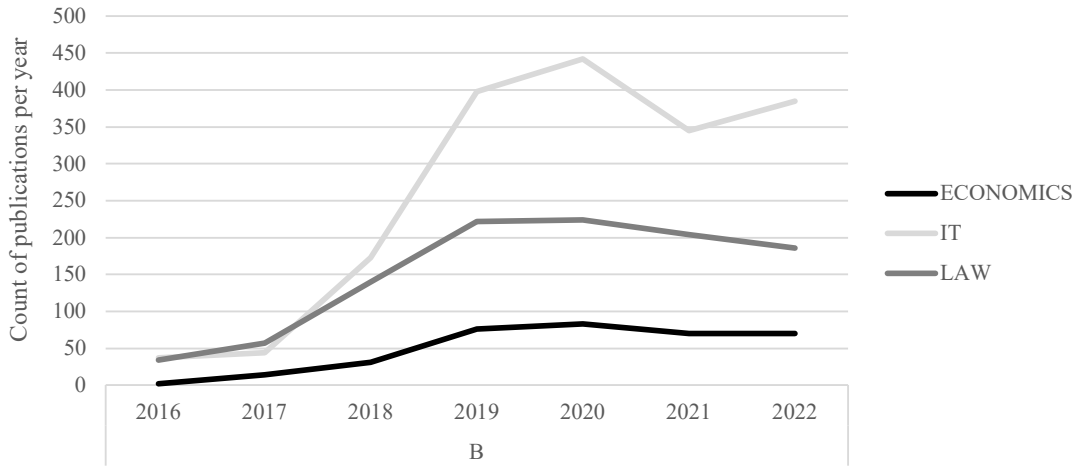

Figure 17: Publications per discipline stage B (n=3,237).

Figure 17 illustrates the evolution of publication rates in the disciplines of law, IT, and economics during stage B from 2016 to 2022. During this period, there was a significant

increase in the number of publications across all disciplines, starting in 2016. In particular, the most significant annual increase across all disciplines occurred between 2018 and 2019.

This spike coincides with the enactment of the General Data Protection Regulation (GDPR) in May 2018 and may reflect the research community's response to the legal implications and challenges introduced by this comprehensive data protection law. The analysis provides valuable insights into the temporal trends in GDPR research output and the research community's responsiveness to major regulatory changes.

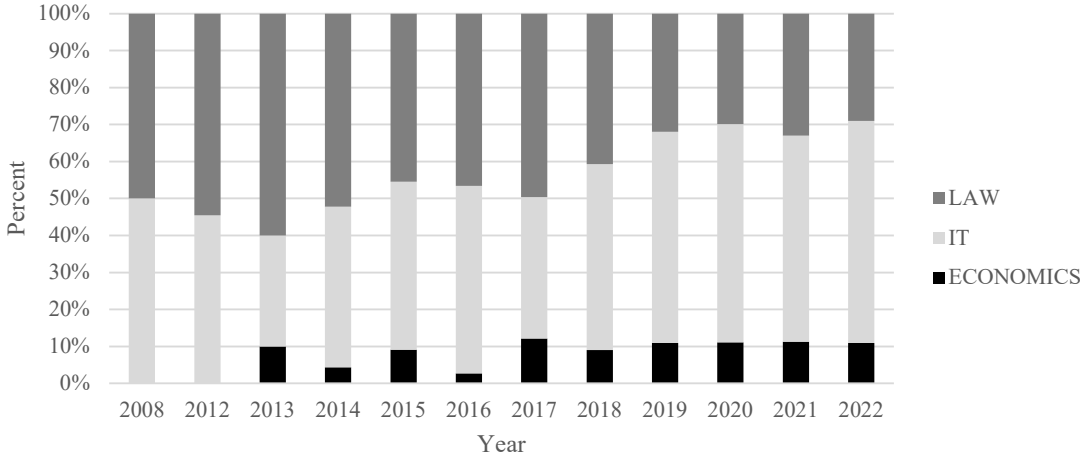

Figure 18: Overall distribution of publications per discipline per year (n=3,315).

Figure 18 shows the annual distribution of publications per discipline from the beginning of the GDPR research until 2022. Initially, the law discipline shows the most significant impact, followed closely by IT. The economics discipline, on the other hand, shows no discernible impact until 2013. Over time, the influence of the law discipline decreases slightly compared to IT, which maintains a consistently high level of importance from the beginning to the end of the observed period. The economics discipline, despite its involvement in GDPR research after 2013, maintains a relatively low role throughout. This figure illustrates the evolving importance of each discipline within the GDPR research landscape and highlights the consistent and enduring importance of IT.

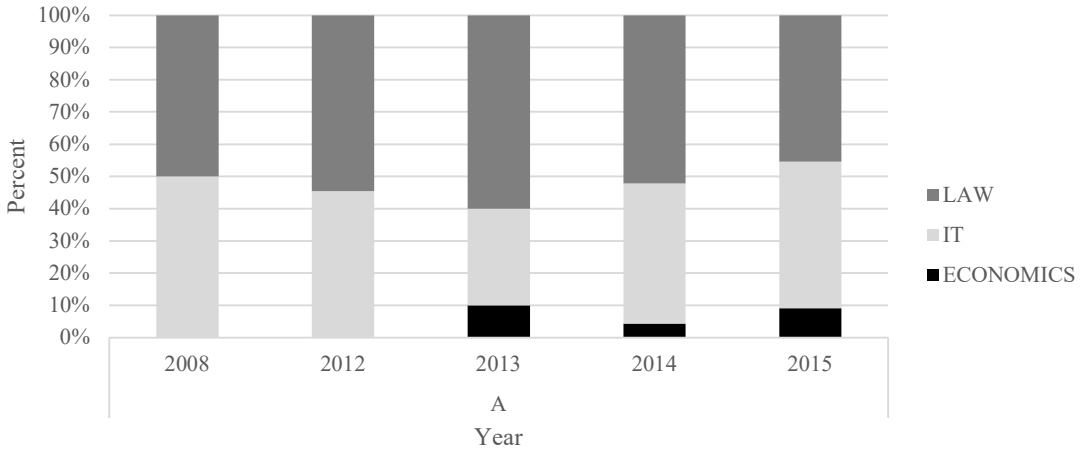

Figure 19: Overall distribution of publications per discipline and year stage A (n=78)

Figure 19 shows the annual distribution of publications per discipline during stage A up until the end of 2015. In this stage, the law discipline's share rises until 2013, after which it experiences a decline. IT, on the other hand, displays high relevance at the beginning, diminishing by 2013 but then rebounding in 2014. Throughout stage A, the representation of the economics discipline is at most 10% of the total publications, indicating a comparatively minor role in GDPR-related research during this period.

This analysis provides a detailed understanding of the disciplinary dynamics within GDPR research during its early years. It points towards the more prominent roles of law and IT disciplines during this stage.

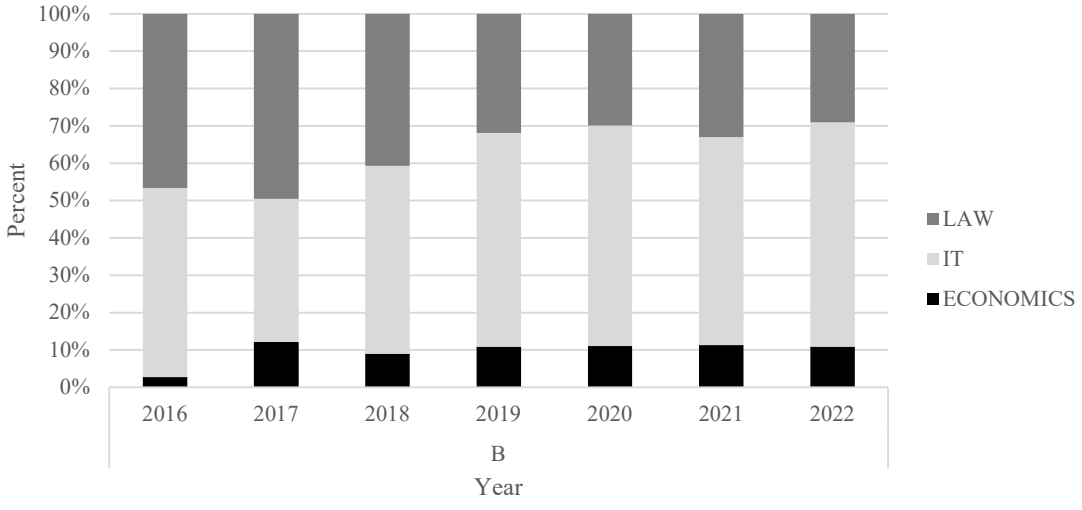

Figure 20: Overall distribution of publications per discipline per year stage B (n=3237).

Figure 20 shows the annual distribution of publications across the law, IT, and economics disciplines during stage B from 2016 to 2022.

Throughout this period, the representation of the economics discipline remains fairly consistent, hovering around 10%, indicating a stable, albeit minor, role in GDPR-related research. In 2016, the importance of the IT discipline surpassed the 50% mark, while the law discipline fell below 50%. Interestingly, the law discipline rebounded in 2017 with the highest number of publications, reflecting a research response to the GDPR adaptation in 2016. This depiction allows us to grasp the changing dynamics and relative influence of each discipline in GDPR research during this later phase, marked by significant regulatory changes.

Table 24: Top-5 most cited papers per discipline and stage. Stage A for law; stage B for law, IT, and economics (n=366)

No.	Stage/ Phase	Discipline	Authors	Title	Year	Journal	Cited in	Global Citations
1	A	LAW	MANTELEYO A	THE EU PROPOSAL FOR A GENERAL DATA PROTECTION REGULATION AND THE ROOTS OF THE `RIGHT TO BE FORGOTTEN'	2013	COMPUTER LAW \& SECURITY REVIEW	LAW	68
2	A	LAW	DE HERT P;PAPAKONSTANTINOU V	THE PROPOSED DATA PROTECTION REGULATION REPLACING DIRECTIVE 95/46/EC: A SOUND SYSTEM FOR THE PROTECTION OF INDIVIDUALS	2012	COMPUTER LAW \& SECURITY REVIEW	LAW	50
3	A	LAW	VICTOR JM	THE EU GENERAL DATA PROTECTION REGULATION: TOWARD A PROPERTY REGIME FOR PROTECTING DATA PRIVACY	2013	YALE LAW JOURNAL	LAW	50
4	A	LAW	MANTELERO A	THE FUTURE OF CONSUMER DATA PROTECTION IN THE EU RE-THINKING THE "NOTICE AND CONSENT" PARADIGM IN THE NEW ERA OF PREDICTIVE ANALYTICS	2014	COMPUTER LAW \& SECURITY REVIEW	LAW	46
5	A	LAW	CUMBLEY R;CHURCH P	IS "BIG DATA" CREEPY?	2013	COMPUTER LAW \& SECURITY REVIEW	LAW	45
1	B	LAW	WACHTER S;MITTELSTADT B;FLORIDI L	WHY A RIGHT TO EXPLANATION OF AUTOMATED DECISION-MAKING DOES NOT EXIST IN THE GENERAL DATA PROTECTION REGULATION	2017	INTERNATIONAL DATA PRIVACY LAW	LAW	354
2	B	LAW	SELBST AD;BAROCAS S	THE INTUITIVE APPEAL OF EXPLAINABLE MACHINES	2018	FORDHAM LAW REVIEW	LAW	157
3	B	LAW	TIKKINEN-PIRI C;ROHUNEN A;MARKKULA J	EU GENERAL DATA PROTECTION REGULATION: CHANGES AND IMPLICATIONS FOR PERSONAL DATA COLLECTING COMPANIES	2018	COMPUTER LAW \& SECURITY REVIEW	LAW	124
4	B	LAW	WACHTER S	NORMATIVE CHALLENGES OF IDENTIFICATION IN THE INTERNET OF THINGS: PRIVACY, PROFILING, DISCRIMINATION, AND THE GDPR	2018	COMPUTER LAW \& SECURITY REVIEW	LAW	87
5	B	LAW	DE HERT P;PAPAKONSTANTINOU V;MALGIERI G;BESLAY L;SANCHEZ I	THE RIGHT TO DATA PORTABILITY IN THE GDPR: TOWARDS USER-CENTRIC INTEROPERABILITY OF DIGITAL SERVICES	2018	COMPUTER LAW \& SECURITY REVIEW	LAW	86

No.	Stage/ Phase	Discipline	Authors	Title	Year	Journal	Cited in	Global Citations
1	B	IT	GOODMAN B;FLAXMAN S	EUROPEAN UNION REGULATIONS ON ALGORITHMIC DECISION MAKING AND A "RIGHT TO EXPLANATION"	2017	AI MAGAZINE	IT	661
2	B	IT	BERNAL BERNABE J;LUIS CANOVAS J;HERNANDEZ-RAMOS JL;TORRES MORENO R;SKARMETA A	PRIVACY-PRESERVING SOLUTIONS FOR BLOCKCHAIN: REVIEW AND CHALLENGES	2019	IEEE ACCESS	IT	112
3	B	IT	KREUTER F;HAAS GC;KEUSCH F;BAEHR S;TRAPPMANN M	COLLECTING SURVEY AND SMARTPHONE SENSOR DATA WITH AN APP: OPPORTUNITIES AND CHALLENGES AROUND PRIVACY AND INFORMED CONSENT	2020	SOCIAL SCIENCE COMPUTER REVIEW	IT	54
4	B	IT	TAMBURRI DA	DESIGN PRINCIPLES FOR THE GENERAL DATA PROTECTION REGULATION (GDPR): A FORMAL CONCEPT ANALYSIS AND ITS EVALUATION	2020	INFORMATION SYSTEMS	IT	37
5	B	IT	CAMPANILE L;IACONO M;MARULLI M	DESIGNING A GDPR COMPLIANT BLOCKCHAIN-BASED IOV DISTRIBUTED INFORMATION TRACKING SYSTEM	2021	INFORMATION PROCESSING \& MANAGEMENT	IT	33
1	B	ECONOMICS	GODDARD M	THE EU GENERAL DATA PROTECTION REGULATION (GDPR): EUROPEAN REGULATION THAT HAS A GLOBAL IMPACT	2017	INTERNATIONAL JOURNAL OF MARKET RESEARCH	ECONOMICS	123
2	B	ECONOMICS	HAMILTON RH;SODEMAN WA	THE QUESTIONS WE ASK: OPPORTUNITIES AND CHALLENGES FOR USING BIG DATA ANALYTICS TO STRATEGICALLY MANAGE HUMAN CAPITAL RESOURCES	2020	BUSINESS HORIZONS	ECONOMICS	54
3	B	ECONOMICS	KUPERBERG M	BLOCKCHAIN-BASED IDENTITY MANAGEMENT: A SURVEY FROM THE ENTERPRISE AND ECOSYSTEM PERSPECTIVE	2020	IEEE TRANSACTIONS ON ENGINEERING MANAGEMENT	ECONOMICS	52
4	B	ECONOMICS	CHOI JP;JEON DS;KIM BC	PRIVACY AND PERSONAL DATA COLLECTION WITH INFORMATION EXTERNALITIES	2019	JOURNAL OF PUBLIC ECONOMICS	ECONOMICS	48
5	B	ECONOMICS	RIEGER A;GUGGENMOS F;LOCKL J;FRIDGEN G;URBACH N	BUILDING A BLOCKCHAIN APPLICATION THAT COMPLIES WITH THE EU GENERAL DATA PROTECTION REGULATION	2019	MIS QUARTERLY EXECUTIVE	ECONOMICS	43

Table 24 presents an analysis of the world's top-cited papers by discipline and stage. It shows the significant papers that have been most influential within each discipline over time. One striking observation from the table is the prevalence of self-citation within disciplines, suggesting that each discipline predominantly cites papers within its own domain. This pattern suggests that while GDPR has had an impact on all three disciplines, the growth in publications per discipline appears to be largely independent, with limited cross-discipline citation interactions. This finding provides an interesting insight into the independent development of GDPR research within the disciplines of law, IT, and economics. It highlights the potential for further exploration of interdisciplinary connections.

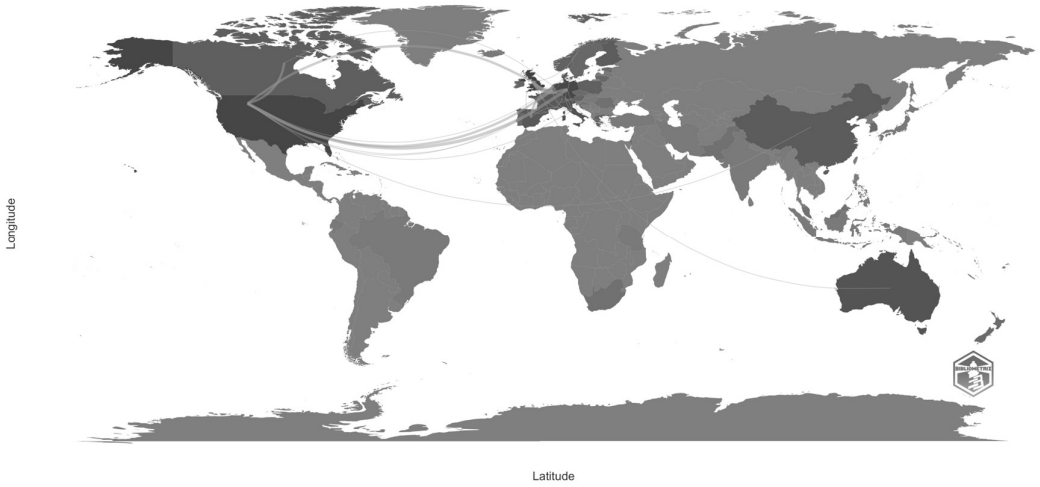

Figure 21: Overall geographical collaboration (n=366).

Figure 21 provides a visualization of the overall geographic collaboration in GDPR-related research. The figure highlights a strong intra-EU collaboration, indicating a significant level of collaboration among researchers within the European Union. This is to be expected, given the origin and primary jurisdiction of the GDPR. In addition, the figure also shows robust research collaboration between the EU and North American countries, particularly the United States and Canada. This finding underscores the global relevance of the GDPR and the importance of transatlantic cooperation in researching its various dimensions. Thus, Figure 21 illuminates the international landscape of GDPR research collaboration, highlighting the central role of the EU and its active engagement with North American partners.

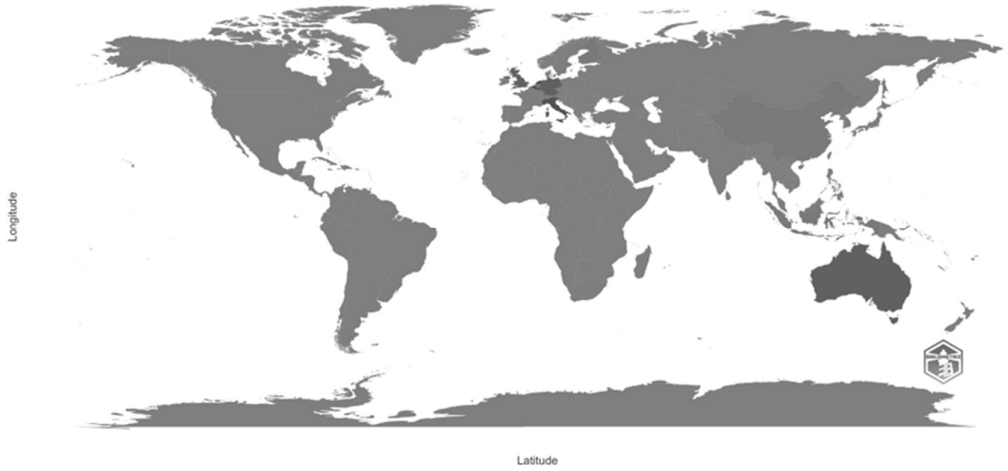

Figure 22: Geographical collaboration stage A (n=18).

Figure 22 shows the geographical collaboration in stage A. It is evident that the earlier stage did not exhibit a significant level of international collaboration. That indicates that the research communities in the different geographical regions worked primarily in isolation during this period, reflecting a period of initial, independent exploration of the GDPR topic.

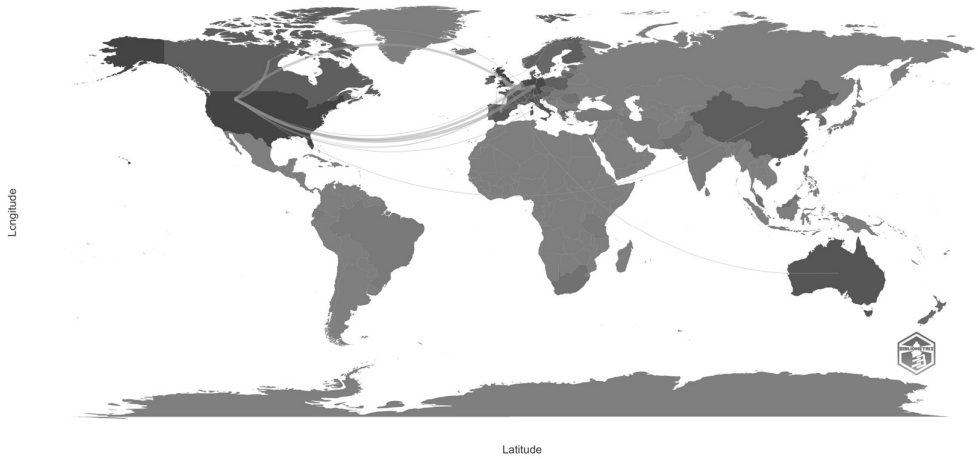

Figure 23: Geographical collaboration stage B (n=348).

Figure 23 presents the geographic cooperation landscape during stage B (2016-2022) of GDPR-related research. During this stage, there is a pronounced pattern of robust collaboration within the European Union. It reflects the primary jurisdiction of the GDPR and its profound impact on research within member states. In addition, the figure reveals an equally strong research collaboration between the EU and its North American counterparts, namely the United States and Canada. This shift from localized exploration in stage A to increased international collaboration in stage B may be due to a growing recognition of the global impact of the GDPR

and the need for cross-border research collaboration to comprehensively address its challenges and opportunities.

Figure 24: Word cloud all disciplines and stages (n=366).

Figure 24 shows a word cloud generated from all disciplines and stages of GDPR-related research, providing a graphical representation of the key terms and themes prevalent in this body of work. In the word cloud, terms such as "data protection," "law," "privacy," and "information" appear prominently, indicating their frequent use across the research corpus. The prevalence of these terms across disciplines and stages underscores the continued centrality of legal compliance and data protection in the GDPR discourse. It provides insight into the key concerns driving research in this area.

Figure 25: Word cloud all disciplines stage B (n=348).

Figure 25 presents a word cloud corresponding to stage B, displaying the predominant terminology across all disciplines during this phase. The representation is comprehensive, featuring a balanced blend of terms from the law, IT, and economics disciplines. This suggests an enriched interdisciplinarity on GDPR-related topics in stage B, underscoring the broader engagement across multiple fields of study during this period.

Figure 26: Word cloud law stage A (n=18).

Figure 27: Word cloud law stage B (n=241).

Figures 26 and 27 illuminate the evolution of topic dominance across the distinct stages A and B within the discipline of law, which uniquely spans both stages. These figures explicitly display a noteworthy shift in topic dominance as research transitions from stage A to B. They underline how new topics emerge and gain prominence in stage B, demonstrating the dynamism and evolution inherent to this area of study. Through these transitions, Figures 26 and 27

effectively capture the shifting landscape of data regulation research within the discipline of law over time.

Figure 28: Word cloud IT stage B (n=45).

Figure 29: Word cloud economics in stage B (n=62).

Figures 28 and 29 show a striking comparison between the key themes dominating the IT and economics disciplines in stage B, as represented by corresponding word clouds. Figure 28 shows the tilt of the IT discipline towards more technical aspects. The prominence of terms such as "information," "security," "Internet," and "big data" underscores the discipline's focus on technological facets and information handling. On the other hand, Figure 29 shows the economics discipline's exploration of GDPR from a distinctly different perspective. Terms such as "impact," "trust," "behavior," "online," and "technology" dominate the discourse. This

underscores the discipline's focus on the impact of data regulation on trust and behavior in online environments, which paints a very different picture from that of IT. The stark contrast between the two figures illustrates the nuanced, discipline-specific approaches to GDPR research in stage B.

Confirmation or Rejection of Hypotheses

The available results will now be examined to see whether they confirm or refute the selected hypotheses (see Table 25).

Table 25: Confirmation or rejection of hypotheses.

Concept	Hypotheses	Confirmed/Rejected
A. Size	A. The publication output on data regulation has grown strongly.	**Confirmed**
	B. The publication volume develops differently per discipline.	**Confirmed**
	C. The growth in publication output is different per discipline.	**Confirmed**
	D. Most cited papers stem from the Law discipline as they are important references for authors in other disciplines.	**Rejected** Most cited papers stem from the IT discipline
B. Time	E. Different publication waves are recognizable. Beginning in the Law discipline, a wave of publications picked up and triggered a growing wave of publications in the disciplines of IT and Economics.	**Confirmed** Different waves are recognizable. Beginning with Law, closely followed by IT, and followed with some delay by the Economics
C. Space	F. The main collaborations take place between the EU and the US. Chinese data privacy researchers do not collaborate that often with researchers from other cultural backgrounds.	**Confirmed** Geographical collaboration takes place to a much greater extent between the EU and the USA than with China
D. Composition	G. Different topics of data regulation research are dominant depending on the discipline and stage.	**Confirmed**

In summary, the comprehensive analysis of the data collected confirms and expands our understanding of the research landscape in relation to the General Data Protection Regulation as follows:

- Hypothesis A & B: The publication output on data regulation has indeed grown significantly, which is corroborated by the trend displayed in our figures. Particularly from 2016 onwards, a robust surge in the number of publications across all disciplines

is noticeable, thereby validating the hypotheses. This escalation in academic interest is reflective of the growing importance of data regulation in the global socio-technological context.

- Hypothesis C: Our research has unveiled a discernible variation in the publication output across different disciplines. In the initial stage, law emerged as the frontrunner with a higher number of publications. The disciplines of IT and economics began contributing significantly only later, suggesting a phased evolution of GDPR-related studies across disciplines. Notably, the IT discipline experienced unprecedented publication growth, underpinning it is escalating relevance in discussions around data regulation.
- Hypothesis D: Contrary to our initial expectation outlined in hypothesis D, our analysis revealed that the most cited papers are primarily from the IT discipline rather than law. This unexpected outcome suggests that IT plays a more influential role in the discourse on data regulation than initially presumed, potentially due to the immediate relevance and applicability of data regulation principles within this field. The law discipline contributed significantly, but our analysis revealed that these citations did not significantly influence authors in other disciplines. This finding points towards discipline-centric growth in GDPR-related research, with each discipline progressing independently.
- Hypothesis E: The existence of distinct publication waves across disciplines is indeed recognizable in our data. The ripple effect commenced in the law discipline, triggered growth in IT, and eventually resonated in the economics discipline. This interdisciplinary cascade confirms the hypothesis and highlights the interconnected yet staggered evolution of research interest in GDPR-related studies across these fields.
- Hypothesis F: Our analysis confirms robust collaborations primarily within the EU and between the EU and the US. The participation of Chinese researchers in data privacy studies, however, appears minimal, suggesting a potential research gap. This lack of diverse international collaborations represents a valuable opportunity for future research efforts.
- Hypothesis G: The data analysis confirms that the dominance of different data regulation research topics varies by discipline and research phase. This finding is exemplified by the word clouds, effectively illustrating the transition in topic dominance across different phases. In addition, there is intriguing variability in the dominance of individual terms within different disciplines. This variability extends to the wide range

of terms used, further highlighting the interdisciplinary and dynamic nature of GDPR-related research.

Conclusion

This comprehensive bibliometric investigation has provided us with a deep and complex understanding of the ever-changing landscape of data regulation research, spanning disciplines such as law, IT, and economics. Our study strongly confirms most of our initial hypotheses. A key finding is a pronounced increase in publication output on data regulation, indicating a significant escalation of scholarly interest in the field. Distinct growth trajectories were observed in each discipline, shedding light on their individual development and research progress. In particular, the discipline of law led the first wave of publications, subsequently stimulating interest in IT and economics, thus confirming Hypothesis E. An unanticipated observation contradicted Hypothesis D; the most cited papers were predominantly from the IT discipline rather than from law. This discrepancy suggests that the role of IT in the data regulation dialogue is crucial, possibly due to the direct relevance and applicability of data regulation principles in this sector.

In addition, our analysis revealed a discernible shift in the prevalence of research topics from stage A to stage B within each discipline. This transition embodies the dynamic nature of data regulation research and the continuous evolution of focus areas in line with the changing regulatory and technological landscape. Our study reveals nuanced insights into the GDPR-related research milieu, underscoring the essential function of interdisciplinary collaboration, highlighting the differential growth across disciplines, and emphasizing the importance of cultivating broader international collaborations to address the multifaceted challenges of data regulation comprehensively. In addition, our research sheds light on the patterns of research collaborations. The most important collaborations were identified within the EU and between the EU and the US, indicating significant transatlantic cooperation. However, a lower frequency of collaboration with Chinese privacy researchers was found, suggesting the existence of potential cultural or systemic barriers that merit further exploration.

Practical Implications

The interdisciplinary lag means that regulatory changes that are first thoroughly analyzed in the legal community may impact IT and economics practices later. This understanding can be used by practitioners, particularly those in IT and business, to prepare for the practical implications of new regulations. For example, IT professionals could anticipate potential changes in data

management protocols, while businesses could prepare for shifts in regulatory compliance requirements. This could lead to a more synchronized response across disciplines, increasing the overall effectiveness of applied research and its subsequent implementation in practice and thus bridging the gap between academic research and practical application.

Regulatory Intelligence for Data Management

Data privacy is relevant to organizations at several levels. To remain competitive, it is essential that management not only implements existing privacy legislation but also prepares for future regulations, especially stricter ones. For this reason, Regulatory Intelligence (RI) has become a key function in the pharmaceutical industry (Ojha 2013). RI professionals analyze and interpret information from various sources to develop a regulatory strategy (Ojha 2013). This information is evaluated regarding its relevance and impact on the company, products, and projects. With their help, impact analyses can be conducted, and compliance projects can be initiated as part of an RC process. This ensures that new requirements are implemented in a timely manner. In addition, these requirements can have an impact on business processes, product development, and approval. This RI/RC process is already known in the pharmaceutical industry (Huddle & Messmer, 2019) and is increasingly being used in medical technology. It, therefore, makes sense to implement RI as early as the R&D phase (Schueler & Ostler 2016).

Theoretical Implications

This study contributes significantly to the scholarly understanding of the evolving research paradigm in data regulation. It highlights the need for interdisciplinary collaboration and dialogue, as the intersection of law, IT, and economics provides a comprehensive view of the complexities of data regulation.

Time lag: The observed time lag between disciplines underscores the different speeds at which disciplines respond to regulatory changes, such as GDPR. Law is naturally the most responsive, given its direct relevance. This finding also suggests a potential for predictive forecasting. A spike in research output in the discipline of law could serve as a harbinger for subsequent waves of intensified research in IT and economics. Scholars in these latter disciplines could therefore gain valuable insights and anticipate future trends by closely monitoring research trajectories in law.

Interdisciplinary Dialogue: The different patterns of publication and topic focus in each discipline suggest that there may be a need for more interdisciplinary dialogue and

understanding. Future research could explore methods to encourage and facilitate cross-disciplinary insights and innovation.

Citation Trends: Given that the most cited papers are predominantly from the IT discipline, it may be worthwhile to investigate the underlying factors that contribute to these citation trends. For example, future studies could examine whether the nature of the topic, the accessibility of the content, or the prominence of the authors or publication venue influence citation frequency.

Barriers to collaboration: The limited collaboration with Chinese researchers provides an opportunity to further investigate the potential barriers to international collaboration. These barriers could be cultural, linguistic, political, or due to differences in research and publication practices.

Topic Evolution: A detailed examination of the temporal evolution of dominant research topics could provide insights into the triggers and effects of these shifts. This could include examining how changes in laws, regulations, societal norms, or technology influence the evolution of research topics.

While our research has provided valuable insights into data regulation across multiple disciplines, it has limitations.

Limitations

First, the paper is based on bibliometric analysis, which inherently depends on the databases' quality and completeness. Our study used two data samples: the large database and the small database. Despite their considerable size, these data sets may only encapsulate some relevant literature on the subject, particularly from non-English language sources or from sources not indexed in Web of Science. In addition, our Large Database could not be thoroughly analyzed using Bibliometrix due to missing parameters, which could limit the comprehensiveness of our findings.

Second, the classification of publications into the disciplines of law, IT, and economics may only partially reflect the interdisciplinary nature of some research, or aspects of all three disciplines may be found in research that was later assigned to only one of the three categories. The inherent challenge of categorizing multidisciplinary research may have led to some misclassifications, affecting the analysis of growth in each discipline.

Third, the word cloud analysis, which provided insights into the dominant themes in each stage and discipline, relies heavily on accurately extracting and categorizing keywords. This

methodology may need to be more balanced with the nuances of the research themes or overlook implicit but important themes.

Future research could overcome these limitations by incorporating additional databases and improving data classification. Despite these limitations, our study provides a solid foundation for understanding the evolving landscape of data regulation research.

References

Acquisti, A., John, L.K., Loewenstein, G. (2013). What Is Privacy Worth? *The Journal of Legal Studies 42(*2):249–274.

Adjerid, I., Peer, E., Acquisti, A. (2018). Beyond the privacy paradox: objective versus relative risk in privacy decision making. *MIS Quarterly* 42(2):465–488. https://doi.org/10.25300/MISQ/2018/14316

Aguinis, H., Ramani, R. S., & Alabduljader, N. (2018). What you see is what you get? Enhancing methodological transparency in management research. *Academy of Management Annals, 12*(1), 83-110.

Aguirre, E., Mahr, D., Grewel, D., Ruyter, K.D., Wetzels M. (2015). Unraveling the personalization paradox: The effect of information collection and trust-building strategies on online advertisement effectiveness. *Journal of Retail 91*(1):34–59. https://doi.org/10.1016/j.jretai.2014.09.005.

Ahrens, J.-P., Isaak A. J., Istipliler B., Steininger, D.M. (2019). The star citizen phenomenon & the 'ultimate dream management' technique in crowdfunding. In: *Proceedings of the 40th international conference on information systems (ICIS)*, Munich, pp 1–9.

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-t.

Ajzen, I., and Fishbein, M. (1980). *Understanding Attitudes and Predicting Social Behavior*. Englewood Cliffs, NJ: Prentice- Hall.

Akil, M., Islami, L., Fischer-Hübner, S., Martucci, L. A., & Zuccato, A. (2020). Privacy-preserving identifiers for IoT: A systematic literature review. *IEEE Access, 8*, 168470-168485.

Akter, S. and Wamba, S.F.(2016). Big data analytics in E-commerce: a systematic review and agenda for future research. *Electronic Markets, 26*, pp.173-194.

Altman, I. (1975). *The environment and social behavior: Privacy, personal space, territory and crowding*. Monterey, C.A.: Brooks/Cole.

Andrews, K. R., Silk, K. S., and Eneli, I. U. (2010). Parents as health promoters: a theory of planned behavior perspective on the prevention of childhood obesity. *Journal of health communication, 15*(1), 95–107. https://doi.org/10.1080/10810730903460567.

Angst, and Agarwal. (2009). Adoption of electronic health records in the presence of privacy concerns: The elaboration likelihood model and individual persuasion. *MIS Quarterly, 33*(2), 339–370. https://doi.org/10.2307/20650295.

Anshari, M., Almunawar, M.N., Lim, S.A. and Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics, 15*(2), pp.94-101.

Aria, M., & Cuccurullo, C. (2017). bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics, 11(4)*, 959-975.

Armitage, C. J., Norman, P., and Conner, M. (2002). Can the theory of planned behaviour mediate the effects of age, gender and multidimensional health locus of control? *British Journal of Health Psychology, 7*(3), 299–316. https://doi.org/10.1348/135910702760213698.

Audretsch, David & Belitski, Maksim. (2017). Entrepreneurial ecosystems in cities: establishing the framework conditions. *The Journal of Technology Transfer*. 42. 10.1007/s10961-016-9473-8.

Baek, Y. M. (2014). Solving the privacy paradox: A counter-argument experimental approach. *Computers in Human Behavior, 38*, 33–42. https://doi.org/10.1016/j.chb.2014.05.006.

Baines, T. & Lightfoot, H. W. (2014). Servitization of the manufacturing firm. International *Journal of Operations & Production Management*, Vol. 34(1), pp. 2-35.

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday, 11*(9). https://doi.org/10.5210/fm.v11i9.1394.

Barney J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management. 17*(1):99-120. doi:10.1177/014920639101700108.

Barth, S., and de Jong, M. D. T. (2017). The privacy paradox – Investigating discrepancies between expressed privacy concerns and actual online behavior – A systematic literature review. *Telematics and Informatics, 34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013.

Barton, D. and Court, D. (2012). Making advanced analytics work for you. *Harvard business review, 90*(10), pp.78-83.

Bauer, M. and Leker, J. (2013). Exploration and exploitation in product and process innovation in the chemical industry. *R&D Management, 43*(3), pp.196-212.

Beck, R., Avital, M., Rossi, M. & Thatcher, J. B., (2017). Blockchain Technology in Business and Information Systems Research. *Business & Information Systems Engineering*: Vol. 59, No. 6. Springer. (S. 381-384). DOI: 10.1007/s12599-017-0505-1.

Becker, M.; Buchkremer, R. (2018). Implementierung einer Regulatory Technology Lösung bei Finanzinstituten unter Berücksichtigung agiler Vorgehensmodelle. In (Mikusz, M.;

Volland, A.; Engstler, M., Hrsg.): *Projektmanagement und Vorgehensmodelle 2018 - Der Einfluss der Digitalisierung auf Projektmanagementmethoden und Entwicklungsprozesse*. Köllen Druck+Verlag GmbH, Bonn, S. 125-134.

Bélanger, F., and Crossler, R. E. (2011). Privacy in the digital age: A review of information privacy research in information systems. *MIS Quarterly, 35*(4), 1017–1041. https://doi.org/10.2307/41409971.

Bélanger, F., Hiller, J. and Smith, W. J. (2002). Trustworthiness in electronic commerce: The role of privacy, security, and site attributes, *Journal of Strategic Information Systems*, 1(3/4), 245-270.

Bennett, C.J. (1992). *Regulating privacy: data protection and public policy in Europe and the United States*, Cornell University Press: New York.

Ben-Shahar, O. and Schneider, C.E. (2014). More than you wanted to know. In *More Than You Wanted to Know.* Princeton University Press.

Berger, Elisabeth & Von Briel, Frederik & Davidsson, Per & Kuckertz, Andreas. (2019). Digital or not – The future of entrepreneurship and innovation. *Journal of Business Research*. 125. 10.1016/j.jbusres.2019.12.020.

Berman, S.J. (2012). Digital transformation: opportunities to create new business models. *Strategy & leadership, 40*(2), pp.16-24.

Biega, A.J. and Finck, M. (2021). Reviving purpose limitation and data minimisation in data-driven systems. *arXiv preprint arXiv:2101.06203*.

Block, J. H., Fisch, C. O., & Van Praag, M. (2018). Quantity and Quality of Jobs by Entrepreneurial Firms. Oxford Review of Economic Policy, 34(4), 565–583. https://doi.org/10.1093/oxrep/gry016.

Bonnefon, J. F., Shariff, A., and Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science, 352*(6293), 1573–1576. https://doi.org/10.1126/science.aaf2654.

Bräunlich, K., Dienlin, T., Eichenhofer, J., Helm, P., Trepte, S., Grimm, R., Seubert, S., & Gusy, C. (2021). Linking loose ends: An interdisciplinary privacy and communication model. *New Media & Society, 23*(6), 1443-1464.

Buchholz, S.; Wirnsperger, P. J.; Wolff, D. (2016). *Zeitgemäßer Datenschutz in der datengetriebenen Wirtschaft - Effektive Umsetzung der EU-Datenschutz-Grundverordnung (DSGVO)*. Deloitte GmbH.

Bygrave, L. A. (2014). *Data Privacy Law: An international perspective*. Oxford University Press.

Bygrave, L.A. (2017). Data protection by design and by default: deciphering the EU's legislative requirements. *Oslo Law Review, 4*(2), pp.105-120.

California Consumer Privacy Act of 2018 ("CCPA"), CAL. CIV. CODE § 1798.198(a) (2018).

Castro, P., Bettencourt, L. (2016). Exploring the predictors and the role of trust and concern in the context of data disclosure to governmental institutions. *Behaviour and Information Technology*. 10.1080/0144929X.2016.1234645.

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., Wirth, R.: CRIPS-DM 1.0 Step by Step Data Mining Guide. CRISP-DM Consortium (2000). Retrived from: http://www.sciepub.com/reference/80444.

Chellappa, R. K., and Sin, R. G. (2005). Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management, 6*(2-3), 181–202. https://doi.org/10.1007/s10799-005-5879-y.

Chen, L., Mislove, A. and Wilson, C. (2015). Peeking beneath the hood of uber. In *Proceedings of the 2015 internet measurement conference* (pp. 495-508).

Chen, Min; Mao, Shiwen; Liu, Yunhao (2014). Big Data: A Survey, in: *Mobile Networks and Applications*, Jahrgang 19, Ausgabe 2, S. 171-209.

Chowdhary, K.R. (2020). *Fundamentals of Artificial Intelligence*. Springer, New Delhi. https://10.1007/978-81-322-3972-7.

Christensen, C. M. (1997). *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Boston, MA: Harvard Business School Press, 1997.

Cleve, J.; Lämmel, U. (2016). *Data Mining*. Berlin, Boston: De Gruyter Oldenbourg. https://doi.org/10.1515/9783110456776-016.

Clough, D. R., and Wu, A. (2020). Artificial Intelligence, data-driven learning, and the decentralized structure of platform ecosystems. *Academy of Management Review*. https://doi.org/10.5465/amr.2020.0222.

Collis, D.J. (1994). Research note: how valuable are organizational capabilities? *Strategic Management Journal. 15 (8)*, 143–152.

Collins, S. E., and Carey, K. B. (2007). The theory of planned behavior as a model of heavy episodic drinking among college students. *Psychology of addictive behaviors : journal of the Society of Psychologists in Addictive Behaviors, 21*(4), 498–507. https://doi.org/10.1037/0893-164X.21.4.498.

Cooper, R.G. and Kleinschmidt, E.J. (1993). Uncovering the keys to new product success. *Engineering Management Review*, *11*(4 S 5), p.18.

Culnan, M.J., and Armstrong, P. (1999). Information Privacy Concerns, Procedural Fairness, and Impersonal Trust: An Empirical Investigation. *Organization Science, 10*, 104-115.

Culnan, M.J. and Milberg, S. (1998). The second exchange: Managing customer information in marketing relationships. *Georgetown McDonough School of Business Research Paper*, (2621796).

Davenport, T. H., and Prusak, L. (2000). *Working knowledge: How organizations manage what they know*. Boston, MA: Harvard Business School Press.

De Luca, L. M., Herhausen, D., Troilo, G., & Rossi, A. (2020). How and when do big data investments pay off? The role of marketing affordances and service innovation. Journal of the Academy of Marketing Science. https://doi.org/10.1007/s11747-020-00739-x.

Del Vecchio, P., Mele, G., Passiante, G., Vrontis, D., & Fanuli, C. (2020). Detecting customers knowledge from social media big data: Toward an integrated methodological framework based on netnography and business analytics. Journal of Knowledge Management, 24(4), 799–821.

Deloitte Touche Tohmatsu Limited (2019). *Künstliche Intelligenz im Compliance-Umfeld von Banken & Co*. Retrieved from: https://www2.deloitte.com/content/dam/Deloitte/de/Documents/risk/Whitepaper-K%C3%BCnstliche-Intelligenz-im-Compliance-Umfeld-von-Banken-und-Co.pdf.

Deloitte Touche Tohmatsu Services, Inc. (2020). *Data valuation: Understanding the value of your data assets*. Retrieved from: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Finance/Valuation-Data-Digital.pdf.

Deuker, A. (2010). *Addressing the privacy paradox by expanded privacy awareness – the example of context-aware services*. In: Bezzi, M., Duquenoy, P., Fischer-Hüber, S., Hansen, M., Zhang, G. (Eds.), Privacy and Identity Management for Life. Springer-Verlag.

Dinev T, Hart P (2004) Internet privacy concerns and their antecedents-measurement validity and a regression model. *Behaviour & Information Technology 23*(6):413–422. https://doi.org/10.1080/01449290410001715723

Dinev, T. and Hart, P. (2006). An Extended Privacy Calculus Model for E-Commerce Transactions. *Information Systems Research. 17*. 61-80. 10.1287/isre.1060.0080.

Dinter, B. and Krämer, J. (2018). Data-driven innovations in electronic markets. *Electronic Markets*, *28*, pp.403-405.

Doyle, G. (2018). Television and the development of the data economy: Data analysis, power and the public interest. *International Journal of Digital Television, 9*(1), 53–68. https://doi.org/10.1386/jdtv.9.1.53_1.

Draper, N.A. and Turow, J. (2019). The corporate cultivation of digital resignation. *New media & society*, *21*(8), pp.1824-1839.

Ducato, R. (2020). Data protection, scientific research, and the role of information. Computer *Law & Security Review, Vol. 37*, 1-16.

Duncan Alan D., Jones Lydia C. (2020). *Applied Infonomics: How to Measure the Net Value of Your Information Assets*. Gartner 2020, ID G00463621.

Eisenhardt, K. M., and Martin, J. A. (2000). Dynamic capabilities: What are they?. *Strategic Management Journal, 21*(10-11), 1105-1121.

El-Amir, H. and Hamdy, M (2020). *Deep Learning Pipeline. Building a Deep Learning Model with TensorFlow*. New York: Apress. https://doi.org/10.1007/978-1-4842-5349-6.

Ellegaard, O. & Wallin, J. A. (2015). The bibliometric analysis of scholarly production How great is the impact?. *Scientometrics, Vol. 105*(3), 1809-1831.

Esposito, F. (2022). The GDPR enshrines the right to the impersonal price. *Computer Law & Security Review, 45*, 105660.

EU Agency for Cybersecurity ENISA. (2015) *Privacy and Data Protection by Design*. Available at: https://www.enisa.europa.eu/publications/privacy-and-data-protection-by-design

EU Data Protection Board. (2019). *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*. Version 2.0. Retrieved from https://edpb.europa.eu/sites/default/files /files/file1/edpb_guidelines-art_6-1-b-adopted_after_public_consultation_en.pdf

EU Directive 95/46: Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, OJ 1995 L 281/31

EU General Data Protection Regulation (GDPR): Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 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 (General Data Protection Regulation), OJ 2016 L 119/1.

Fairclough N (1995). *Media Discourse*. Edward Arnold: London.

Felzmann, H., Villaronga, E.F., Lutz, C. and Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society, 6*(1), p. 1–14.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: Impact on firm performance. *Management Decision, 57*(8), 1923–1936.

Fichman, R.G., Dos Santos, B.L. and Zheng, Z. (2014). Digital innovation as a fundamental and powerful concept in the information systems curriculum. *MIS quarterly*, *38*(2), pp.329-343.

Forgó, N., Hänold, S. and Schütze, B. (2017). The principle of purpose limitation and big data. *New technology, big data and the law*, pp.17-42.

Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D. (2015). How ‘‘big data” can make big impact: findings from a systematic review and a longitudinal case study. *International Journal of Production Economics. 165*, 234–246.

Freiknecht, Jonas (2018). *Big Data in der Praxis: Lösungen mit Hadoop, Spark, HBase und Hive. Daten speichern, aufbereiten, visualisieren*. 2. Auflage, München: Carl Hanser Verlag.

Frye, D.; Palfai, T.; Zelazo, D.P. (1995). *Theory of mind and rule-based reasoning*. In (Cognitive Development, Hrsg.): Cognitive Development, Volume 10, Issue 4. S. 483-527.

Gao, P., Wu, W., & Yang, Y. (2022). Discovering themes and trends in digital transformation and innovation research. *Journal of Theoretical and Applied Electronic Commerce Research, 17*(3), 1162-1184

Gasser, U. (2016). Recoding privacy Law: Reflections on the future relationship among Law, technology, and privacy. *Harvard Law Review, F. 130*, 61.

Gawer, A. (2014). Bridging differing perspectives on technological platforms: Toward an integrative framework. *Research Policy*, 1239-1249.

Ghani, N.A., Hamid, S. and Udzir, N.I. (2016). Big data and data protection: Issues with purpose limitation principle. *International Journal of Advances in Software Computing & Its Applications*, *8*(3).

Gharib, M. (2022). Privacy and informational self-determination through informed consent: the way forward. In *Computer Security. ESORICS 2021 International Workshops: CyberICPS, SECPRE, ADIoT, SPOSE, CPS4CIP, and CDT&SECOMANE,*

Darmstadt, Germany, October 4–8, 2021, Revised Selected Papers (pp. 171-184). Cham: Springer International Publishing.

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences. *Information Processing & Management, 35*(1), 31-44.

Gödert, W. (2010). *Semantische Wissensrepräsentation und Interoperabilität*. In (Deutsche Gesellschaft für Informationswissenschaft und Informationspraxis e. V., Hrsg.): Information - Wissenschaft & Praxis, 61. Jahrgang, Nr. 1. S. 5-28.

Goicovici, J. (2019). Consumer's Consent to the Processing of Personal Data in Business to Consumer Contracts-The Requirement of Granular Consent. *Law Series Annals WU Timisoara*, p.7.

Goldfarb, A and Tucker, C. (2011). Privacy and Innovation. *Innovation Policy and the Economy*. 12. 10.1086/663156.

Good, N., Dhamija, R., Grossklags, J., Thaw, D., Aronowitz, S., Mulligan, D. and Konstan, J. (2005). Stopping spyware at the gate: a user study of privacy, notice and spyware. In *Proceedings of the 2005 symposium on Usable privacy and security* (pp. 43-52).

Greenleaf, G. (2019). Global Data Privacy Laws. 132 National Laws and Many Bills (February 8, 2019). *Privacy Laws and Business International Report*, 14-18. Retrieved from SSRN: https://ssrn.com/abstract=3381593.

Gregory, R. W., Henfridsson, O., Kaganer, E., and Kyriakou, S. H. (2020). The role of artificial intelligence and data network effects for creating user value. *Academy of Management Review, 46*(3), 534–551. https://doi.org/10.5465/amr.2019.0178.

Gumus S., Bellibas M. S., Esen M., Gumus E. (2018). A systematic review of studies on leadership models in educational research from 1980 to 2014. *Educational Management Administration & Leadership, 46*, 25–48.

Hackett, D. (2016). *Big Data in Life Insurance*. Retrieved from: https://www.mlc.com.au/content/dam/mlc/documents/pdf/media-centre/big-data-report.pdf.

Haefner, N., Wincent, J., Parida, V. and Gassmann, O. (2021). Artificial intelligence and innovation management: A review, framework, and research agenda☆. *Technological Forecasting and Social Change*, *162*, p.120392.

Hahn, I. (2021). Purpose Limitation in the Time of Data Power: Is There a Way Forward?. *Eur. Data Prot. L. Rev.*, *7*, p.31.

Hallinger, P., & Kovačević, J. (2019). A bibliometric review of research on educational administration: Science mapping the literature, 1960 to 2018. *Review of Educational Research, 89*(3), 335-369.

He, W. (2012). *In the name of justice: Striving for the rule of law in China*. Brookings Institution Press.

He, X. (2012). The party's leadership as a living constitution in China. *Hong Kong LJ 42*, 73. https://heinonline.org/HOL/LandingPage?handle=hein.journals/honkon42anddiv=9anid=andpage=.

Hellmann, R. (2018). *IT-Sicherheit: Eine Einführung*, Berlin, Boston: De Gruyter Oldenbourg, 2018. https://doi.org/10.1515/9783110494853.

Henfridsson, O., Mathiassen, L. and Svahn, F. (2014). Managing technological change in the digital age: the role of architectural frames. *Journal of Information Technology*, *29*(1), pp.27-43.

Hinde, S. (1998). Privacy and security—The drivers for growth of E-Commerce. *Computers & Security*, *17*(6), pp.475-478.

Hoffmann, C. P., Lutz, C., and Ranzini, G. (2016). Privacy cynicism: A new approach to the privacy paradox. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 10*(4), 7. https://doi.org/10.5817/cp2016-4-7.

Hui, K.L. and Chau, P.Y. (2002). Classifying digital products. *Communications of the ACM*, *45*(6), pp.73-79.

IBM Security & Ponemon Institute (2019). Annual Study: The Cost of a Data Breach Report, Retrieved from https://www.ibm.com/security/data-breach.

ICO (2020). The UK Information Commissioner's Office (ICO), *Enforcement Information: British Airways*, Retrieved from: https://ico.org.uk/action-weve-taken/enforcement/british-airways/.

Isaak, J. and Hanna, M. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51, pp. 56–59.

Kahneman D (2003). A perspective on judgement and choice: mapping bounded rationality. *American Psychologist 58*(9):697–720. https://doi.org/10.1037/0003-066X.58.9.697

Kamara, I. and De Hert, P. (2018). Understanding the balancing act behind the legitimate interest of the controller ground: A pragmatic approach. *Brussels Privacy Hub*, *4*(12).

Karnouskos, S., and Kerschbaum, F. (2017). Privacy and integrity considerations in hyperconnected autonomous vehicles. *Proceedings of the IEEE, 106*(1), 160–170. https://doi.org/10.1109/jproc.2017.2725339.

Kelley, P.G., Bresee, J., Cranor, L.F. and Reeder, R.W. (2009). A" nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (pp. 1-12).

Kennedy, H. (2018). Living with data: Aligning data studies and data activism through a focus on everyday experiences of datafication, *Krisis: Journal for Contemporary Philosophy*. Available at http://krisis.eu/living-with-data/.

Khatri, V., and Brown, C. V. (2010). Designing data governance. *Communications of the ACM, 53*(1), 148–152. https://doi.org/10.1145/1629175.1629210.

Kitsiou, A., Tzortzaki, E., Kalloniatis, C., & Gritzalis, S. (2021). Identifying privacy related requirements for the design of self-adaptive privacy protections schemes in social networks. *Future Internet, 13*(2), 23.

Kittel, K. (2013). *Agilität von Geschäftsprozessen trotz Compliance*. In (Wirtschaftsinformatik Proceedings, Hrsg.): Wirtschaftsinformatiok Proceedings 2013. S. 967-981.

Knape, T., Hufnagl, P., & Rasche, C. (2020). Dashboardconsulting im Gesundheitswesen – Digitalisierungsoptionen und Anwendungsfelder, in: Pfannstiel, M., Rasche, C., Braun von Reinersdorff, A., Knoblach, B. & Fink, D. (Hrsg.): *Consulting im Gesundheitswesen – Professional Services als Gestaltungsimperative in der Gesundheitswirtschaft*, Wiesbaden, S. 1-27.

Koehler, W., Schultz, C. & Rasche, C. (2022). Data are the fuel for digital entrepreneurship - but what about data privacy? In Handbook of Digital Entrepreneurship, 306–322. https://doi.org/10.4337/9781800373631.00027

Koehler, W.; Schultz, C.; Rasche, C. (2020). *Das 100% Problem im Datenschutz*. G-Forum Konferenz, Karlsruhe. Retrived from: https://www.uni-potsdam.de/fileadmin/projects/professional-services/downloads/dokumente/_Praxis-Beitrag__Das_100_-Problem_Datenschutz_G_2020_sent.pdf.

Koo, M. (2017). A bibliometric analysis of two decades of aromatherapy research. *BMC Research Notes, 10*(1), 46. doi:10.1186/s13104-016-2371-1

Kosta, E. (2013). *Consent in European data protection law*. Martinus Nijhoff Publishers.

Kollmann, T. (2016). *E-Entrepreneurship: Grundlagen der Unternehmensgründung in der Digitalen Wirtschaft*. Wiesbaden: Springer Gabler-Verlag.

Krasnova, H., Spiekermann, S., Koroleva, K., and Hildebrand, T. (2010). Online social networks: Why we disclose. *Journal of Information Technology, 25*(2), 109–125. https://doi.org/10.1057/jit.2010.6.

Krueger, N. F., Reilly, M. D., and Carsrud, A. L. (2000). Competing models of entrepreneurial intentions. *Journal of Business Venturing, 15*(5-6), 411–432. https://doi.org/10.1016/s0883-9026(98)00033-0.

Labadie, C., & Legner, C. (2023). Building data management capabilities to address data protection regulations: Learnings from EU-GDPR. *Journal of Information Technology, 38*(1), 16-44.

Lamot, K., and Paulussen, S. (2020). Six uses of analytics: Digital editors' perceptions of audience analytics in the newsroom. *Journalism Practice, 14*(3), 358–373. https://doi.org/10.1080/17512786.2019.1617043.

Lanier, C., and Saini, A. (2008). Understanding consumer privacy: A review and future directions. *Academy of Marketing Science Review, 12*(2), 1–45.

Lasarov, W., Hoffmann, S. Paradoxes Datenschutzverhalten. *HMD* (2021). https://doi.org/10.1365/s40702-021-00706-2

Laufer, R. S., and Wolfe, M. (1977). Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of Social Issues*, 33(3), 22–42. https://doi.org/10.1111/j.1540-4560.1977.tb01880.x.

Lee, H.; Kweon, E.; Kim, M.; Chai, S. (2017). Does Implementation of Big Data Analytics Improve Firms' Market Value? Investors' Reaction in Stock Market. *Sustainability 2017*, 9, 978. https://doi.org/10.3390/su9060978.

Legner, C., Eymann, T., Hess, T., Matt, C., Böhmann, T., Drews, P., Mädche, A., Urbach, N. and Ahlemann, F. (2017). Digitalization: opportunity and challenge for the business and information systems engineering community. *Business & information systems engineering*, *59*, pp.301-308.

Lenhard, J., Fritsch, L. & Herold, S. (2017). *A literature study on privacy patterns research, 2017*. 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA), 194-201, doi: 10.1109/SEAA.2017.28.

Li, W.C.Y., Nirei, M., & Yamana, K. (2018). Value of Data: There's No Such Thing As A Free Lunch in the Digital Economy, Presentation at the *Sixth IMF Statistical Forum*, Washington DC, November 2018. Retrieved from: https://unstats.un.org/unsd /nationalaccount/aeg/2018/M12_3c2_Data_SNA_asset_boundary.pdf.

Liao, C., Chen, J.L. and Yen, D.C. (2007). Theory of planning behavior (TPB) and customer satisfaction in the continued use of e-service: An integrated model. *Computers in human behavior*, *23*(6), pp.2804-2822.

Libert, B., (2013). *Why boards must embrace big data*. NACD Director. (September/October), 26–30.

Liu, B., Pavlou, P.A. and Cheng, X. (2022). Achieving a balance between privacy protection and data collection: A field experimental examination of a theory-driven information technology solution. *Information Systems Research*, *33*(1), pp.203-223.

Loebbecke, C. and Picot, A. (2015). Reflections on societal and business model transformation arising from digitization and big data analytics: A research agenda. *The Journal of Strategic Information Systems*, *24*(3), pp.149-157.

Lokhande, S. A., & Khare, N. (2015). An Outlook on Big Data and Big Data Analytics. *International Journal of Computer Applications* 124(11), pp. 37-41.

Lupton, D. (2019). Data selves: *More-than-human perspectives*. Cambridge: Polity Press.

Lupton, D. and Michael, M. (2017). Depends on who's got the data: Public understandings of personal digital dataveillance, *Surveillance & Society*, 15, pp. 254–268.

Lycett, M. (2013). 'Datafication': making sense of (big) data in a complex world. *European Journal of Information Systems. 22*, 381–386.

Lynskey, O. (2015). *The foundations of EU data protection law*. Oxford University Press.

Lyytinen, K., Yoo, Y. and Boland Jr, R.J. (2016). Digital product innovation within four classes of innovation networks. *Information Systems Journal*, *26*(1), pp.47-75.

Malgieri, G. and Niklas, J. (2020). Vulnerable data subjects. *Computer Law & Security Review*, *37*, p.105415.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., Byers, A.H., (2011). *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, San Francisco.

Marchand, D., Kettinger, W., Rollins, J. (2000). Information orientation: people, technology and the bottom line. *Sloan Management Review. 41* (4), 69–80.

Martin, K., Nissenbaum, H. (2016). Measuring privacy: an empirical test using context to expose confounding variables. *The Columbia Science and Technology Law Review 18*:176

Martin, N., Matt, C., Niebel, C. and Blind, K. (2019). How data protection regulation affects startup innovation. *Information Systems Frontiers*. 10.1007/s10796-019-09974-2.

Mason, R. O. (1986). Four ethical issues of the information age*, MIS Quarterly, 10*(1), 5-12.

McAfee, A., Brynjolfsson, E. (2008). Investing in the IT that makes a competitive difference. *Harvard Business Review. 86*, 99–107.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. and Barton, D. (2012). Big data: the management revolution. *Harvard business review, 90*, 60-68.

Melumad S., Meyer R. (2020) Full disclosure: how Smartphones enhance consumer self-disclosure. *Journal of Marketing 84*(3):28–45. https://doi.org/10.1177/0022242920912732

Mikalef, P., Pappas, Ilias O., Krogstie, John., Giannakos, M. (2017). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, pp. 1-32. Springer (2017).

Mittelstadt B. D., Allo P., Taddeo M., Wachter S., and Floridi L. (2016). The ethics of algorithms: Mapping the debate. *Big Data and Society 3*, 2, 1–21. https://doi.org/10.1177/2053951716679679.

Mou, J., Cui, Y., & Kurcz, K. (2019). Bibliometric and visualized analysis of research on major e-commerce journals using CiteSpace. J*ournal of Electronic Commerce Research, 20*(4), 219–237.

Nambisan, S., Lyytinen, K., Majchrzak, A., Song, M. (2017). Digital Innovation Management: Reinventing Innovation Management Research in a Digital World. *MIS Quarterly*. 41. 10.25300/MISQ/2017/41:1.03.

Neely, A. (2019). OpenDataScience*. Data Valuation – What is Your Data Worth and How do You Value it?*: Retrieved from https://opendatascience.com/data-valuation-what-isyour-data-worth-and-how-do-you-value-it/ (accessed on October 3, 2020).

Niedermeier, R. and Mpame, M.E. (2019). Processing Personal Data under Article 6 (f) of the GDPR: The Concept of Legitimate Interest. *Int'l J. Data Protection Officer, Privacy Officer & Privacy Couns.*, *3*, p.18.

Nissenbaum, H. (2004). Privacy as contextual integrity. *Wash. L. Rev.*, *79*, p.119.

Nissenbaum, H. (2009). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford, CA: Stanford University Press.

Nonaka, I., Toyama, R. & Konno, N. (2000). SECI, Ba and Leadership: A Unified Model of Dynamic Knowledge-Creation. *Long Range Planning. Vol. 33*, No. 1, S. 5–34.

Norberg PA, Horne DR, Horne DA (2007). The privacy paradox: personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs 41*(1):100–126. https://doi.org/10.1111/j.1745-6606.2006.00070.x.

Obar, J. A., and Oeldorf-Hirsch, A. (2020). The biggest lie on the Internet: Ignoring the privacy policies and terms of service policies of social networking services.

Information, Communication and Society, 23(1), 128–147. https://doi.org/10.1080/1369118x.2018.1486870.

Ojha, Rajashri (2013). *Regulatory Intelligence - need of the hour. Asian Journal of Pharmaceutical Research and Development*. Available online: https://ajprd.com/index.php/journal/article/view/66/61 (accessed on 3/23/23).

O'Leary, R. (2020). *Worldwide Data Privacy Management Software Market Shares*, 2019: OneTrust Dominates the Competition. Framingham: International Data Corporation (IDC).

Oplatka I., Arar K. H. (2017). The research on educational leadership and management in the Arab world since the 1990s: A systematic review. *Review of Education, 5*, 267–307.

Opresnik, D., and Taisch, M. (2015). The value of Big Data in servitization. *International Journal of Production Economics, 165*, 174–184. https://doi.org/10.1016/j.ijpe.2014.12.036.

Paal P. and Pauly D. (2018). *Kommentar zur Datenschutzgrundverordnung und dem Bundesdatenschutzgesetz*. Munich: C.H. Beck.

Parker, G., Van Alstyne, M., & Choudary, S. (2017). *Die Plattform-Revolution: Von Airbnb, Uber, PayPal und co. lernen: Wie neue Plattform-Geschäftsmodelle die Wirtschaft verändern*. Frechen: mitp Verlags GmbH & Co. KG.

Penrose, E. T. (1959). *The Theory of the Growth of the Firm*. JohnWiley, New York.

Peng, D.X., Schroeder, R.C., Shah, R. (2008). Linking routines to operations capabilities: a new perspective. *Journal of Operations Management*. 26 (6), 730–748.

Pernot-Leplay, E. (2020). China's approach on data privacy law: A third way between the US and the EU. *Penn State Journal of Law and International Affairs, 8*, 49.

Peteraf, M.A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal*, 14: 179-191. https://doi.org/10.1002/smj.4250140303.

Personal Information Protection and Electronic Documents Act (PIPEDA), SC 2000, c 5, Retrieved from: https://canlii.ca/t/541b8. Accessed on 2021-10-04

Peters, R.; Nauroth, M. (2019). *Process-Mining Geschäftsprozesse: smart, schnell und einfach*. Springer Gabler, Wiesbaden.

Phelan C., Lampe C., and Resnick P. (2016). It's creepy, but it doesn't bother me. In CHI '16: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose, CA, USA, 5240–5251. https://doi.org/10.1145/2858036.2858381.

Pilkington A., Meredith J. (2009). The evolution of the intellectual structure of operations management 1980–2006: A citation/co-citation analysis. *Journal of Operations Management, 27*, 185-202.

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., and Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879.

Porter, M.E. and Heppelmann, J.E., 2014. How smart, connected products are transforming competition. *Harvard business review*, *92*(11), pp.64-88.

Porter, M. E., and Heppelmann, J. E. (2015). How smart, connected products are transforming companies. *Harvard Business Review, 93*(10), 96–114.

Porter, M. and van der Linde, C. (1995). Toward a new conception of the environment-competitiveness relationship. *The Journal of Economic Perspectives*, 9(4), 97–118.

Prahalad, C.K. and Hamel, G. (1990). The Core Competence of the Corporation. *Harvard Business Review*. 79-91.

PricewaterhouseCoopers LLP (2019). The 2019 Strategy and Digital Auto Report: Time to get real: Opportunities in a transforming market. Retrieved from https://www.strategyand.pwc.com/de/en/industries/automotive/digital-auto-report-2019/digital-auto-report-2019.pdf.

PricewaterhouseCoopers LLP (2019). *Putting a value on data*. Retrieved from: https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf [accessed on Dec 22 2020].

Rainie, L., and Madden, M. (2015). *Americans' privacy strategies post-Snowden*: Pew Research. Retrieved from https://www.pewresearch.org/internet/2015/03/16/americans-privacy-strategies-post-snowden/.

Rajesh, S., Swapna, S., Shylender, P., Reddy, D. (2012). Data as a Service (Daas) in Cloud Computing. *Global Journal of Computer Science and Technology*. Retrieved from https://computerresearch.org/index.php/computer/article/view/286.

Rasche, C. (2020): Nicht-Markt-Strategien im Gesundheitswesen – Wettbewerbsvorteile durch indirektes Management. In: FOR-MED, *Zeitschrift für das Management im Gesundheitswesen*, Ausgabe 04/2020, S. 9-19.

Recchia, G., & Jones, M. N. (2009). More data trumps smarter algorithms: Comparing pointwise mutual information with Latent Semantic analysis. *Behavior Research Methods*, *41*(3), 647–656. https://doi.org/10.3758/brm.41.3.647.

Recker, J. & von Briel, F. (2019). The Future of Digital Entrepreneurship Research: Existing and Emerging Opportunities. *40th International Conference on Information Systems*.

Regan, P. M. (1995). *Legislating privacy: Technology, social values, and public policy*. Univ of North Carolina Press: Chapel Hill.

Reinkemeyer, L. (2020). *Process Mining in Action Principles, Use Cases and Outlook*. Springer Nature Switzerland AG.

Reinsel, D., Gantz, J., and Rydning, J. (2017). Data Age 2025: The evolution of data to life-critical: Don't focus on big data; Focus on the Data That's Big. Framingham: *International Data Corporation* (IDC).

Rippa, P. & Secundo, G. (2019). Digital academic entrepreneurship: The potential of digital technologies on academic entrepreneurship. *Technological Forecasting and Social Change*, 2019, vol. 146, issue C, 900-911.

Roberts, D.L. and Candi, M. (2014). Leveraging social network sites in new product development: Opportunity or hype?. *Journal of Product Innovation Management*, *31*, pp.105-117.

Romeike, F. (2019). RiskNET. Von Im Risikomanagement die Stärken von KI nutzen: https://www.risknet.de/themen/risknews/ im-risikomanagement-die-staerken-von-ki-nutzen/.

Salehan, M. and Kim, D.J. (2016). Predicting the performance of online consumer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, *81*, pp.30-40.

Sandvig, C., Hamilton, K., Karahalios, K. and Langbort, C. (2014). Auditing algorithms: Research methods for detecting discrimination on internet platforms. *Data and discrimination: converting critical concerns into productive inquiry*, *22*(2014), pp.4349-4357.

Sarin, S. and O'Connor, G.C. (2009). First among equals: The effect of team leader characteristics on the internal dynamics of cross-functional product development teams. *Journal of product innovation management*, *26*(2), pp.188-205.

Schermer, B.W. (2011). The limits of privacy in automated profiling and data mining. *Computer Law & Security Review*, *27*(1), pp.45-52.

Schmidt, K. J. (2020). *Datenschutz als Vermögensrecht - Datenschutzrecht als Instrument des Datenhandels*. (M. Cornils, & L. Specht-Riemenschneider, Hrsg.) Wiesbaden: Springer Fachmedien Wiesbaden GmbH.

Schoemaker, P. J., Heaton, S., and Teece, D. (2018). Innovation, dynamic capabilities, and leadership. *California Management Review, 61*(1), 15-42.

Schueler, J., Ostler, T. (2016). Biopharmaceutical Startup's Need of Regulatory Intelligence. *Journal of Commercial Biotechnology, 22*, 6-14.

Schulte-Althoff, M., Fuerstenau, D., & Lee, G. M. (2021). A Scaling Perspective on AI Startups. Proceedings of the 54th Hawaii International Conference on System Sciences (pp. 6515-6524). Honolulu: *Hawaii International Conference on System Sciences (HICSS)*.

Schultz, C. (2021). A Balanced Strategy for Entrepreneurship Education: Engaging Students by Using Multiple Course Modes in a Business Curriculum. *Journal of Management Education,* online first, https://doi.org/10.1177/10525629211017958.

Schultz, C.; Mietzner, D., and Hartmann, F. (2016). Action research as a viable research methodology in entrepreneurship research, in: Berger, E. and Kuckertz, A. (eds.): *Complexity in entrepreneurship, innovation and technology research - Applications of emergent and neglected methods*. Bern: Schweiz: Springer International Publishing, pp. 267-283. https://doi.org/10.1007/978-3-319-27108-8_13.

Seiberth, G., & Gruendinger, W. (2018). *Data-driven Business Models in Connected Cars, Mobility Services and Beyond*. Berlin: BVDW Research, No. 01/18, April 2018.

Shapiro, C. and Varian, H.R. (1998). Versioning: the smart way to. *Harvard business review*, *107*(6), p.107.

Shim, J., French, A., Guo, C. & Jablonski, J. (2015). Big Data and Analytics: Issues, Solutions, and ROI. *Communications of the Association for Information Systems*. 37. 797-810.

Shklovski, I., Mainwaring, S.D., Skúladóttir, H.H., Borgthorsson, H. (2014). Leakiness and creepiness in app space: Perceptions of privacy and mobile app use. CHI 2014: 2347-2356.

Shu, S., & Liu, Y. (2021). Looking back to move forward: A bibliometric analysis of consumer privacy research. *Journal of Theoretical and Applied Electronic Commerce Research, 16*(4), 727-747.

Siddiqa, A., Hashem, I. A. T., Yaqoob, I., Marjani, M., Shamshirband, S., Gani, A., and Nasaruddin, F. (2016). A survey of big data management: Taxonomy and state-of-the-

art. *Journal of Network and Computer Applications, 71*, 151–166. https://doi.org/10.1016/j.jnca.2016.04.008.

Simon, H. (1955). A Behavioral Model of Rational Choice, The Quarterly Journal of Economics, Volume 69 (1): 99–118, https://doi.org/10.2307/1884852

Slotin, J., (2018). What Do We Know About the Value of Data?. *Global Partnership for Sustainable Development Data*. Retrieved from: http://www.data4sdgs.org/news/what-do-we-know-about-value-data.

Smith, H. J., Dinev, T., and Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989–1015. https://doi.org/10.2307/41409970.

Smith, R. E. (2000). Ben Franklin's web site: Privacy and curiosity from Plymouth Rock to the Internet. *Privacy Journal*: Providence, USA .

Soley, A. M., Siegel, J. E., Suo, D., and Sarma, S. E. (2018). Value in vehicles: Economic assessment of automotive data. *Digital Policy, Regulation and Governance, 20*(6), 513–527. https://doi.org/10.1108/dprg-05-2018-0025.

Solove, D.J. (2013). Introduction: Privacy self-management and the consent dilemma. *Harv. L. Rev.*, *126*, p.1880.

Solove, D. J., Schwartz, P. M. (2020). *Information privacy Law*. Aspen Publishing: New York.

Spiekermann, S., Acquisti, A., Böhme, R. and Hui, K.L. (2015). The challenges of personal data markets and privacy. *Electronic markets*, *25*, pp.161-167.

Taddicken, M. (2014). The 'privacy paradox'in the social web: The impact of privacy concerns, individual characteristics, and the perceived social relevance on different forms of self-disclosure. *Journal of computer-mediated communication*, *19*(2), pp.248-273.

Tallon, P. P., and Scannell, R. (2007). Information life cycle management. *Communications of the ACM, 50*(11), 65–69. https://doi.org/10.1145/1297797.1297799.

Tallon, P. P., Ramirez, R. V., and Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems, 30*(3), 141–178. https://doi.org/10.2753/mis0742-1222300306.

Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350. https://doi.org/10.1002/smj.640.

Teece, D. J. (2018). Tesla and the reshaping of the auto industry. *Management and Organization Review, 14*(3), 501–512. https://doi.org/10.1017/mor.2018.33.

Teece, D. J., Pisano, G., and Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(sici)1097-0266(199708)18:7<509::aid-smj882>3.0.co;2-z.

Thanuskodi, S. (2010). Journal of Social Sciences: A bibliometric study. *Journal of Social Sciences, 24*(2), 77-80.

Thompson, D. F., & Walker, C. K. (2015). A descriptive and historical review of bibliometrics with applications to medical sciences. *Pharmacotherapy, 35*(6), 551-559. doi:10.1002/phar.1586

Tonetti, C., & Jones, C. I. (2020). Nonrivalry and the Economics of Data. *American Economic Review* 110(9), S. 2819–2858. DOI: 10.1257/aer.20191330.

Tutton, R. (2017). Wicked futures: Meaning, matter and the sociology of the future', *The Sociological Review*, 65, pp. 478–492.

Utterback, J. (1994). *Mastering the dynamics of innovation: How companies can seize opportunities in the face of technological change*: University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.

van Dijck, J. (2014). Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology, Surveillance & Society, 12, pp. 197–208.

Varadarajan, R., Welden, R.B., Arunachalam, S., Haenlein, M. and Gupta, S. (2022). Digital product innovations for the greater good and digital marketing innovations in communications and channels: Evolution, emerging issues, and future research directions. *International Journal of Research in Marketing*, *39*(2), pp.482-501.

Venkatadri, G., Andreou, A., Liu, Y., Mislove, A., Gummadi, K.P., Loiseau, P. and Goga, O. (2018). Privacy risks with Facebook's PII-based targeting: Auditing a data broker's advertising interface. In *2018 IEEE Symposium on Security and Privacy (SP)* (pp. 89-107). IEEE.

Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems, 28*(2), 118–144. https://doi.org/10.1016/j.jsis.2019.01.003.

Vijayakumaran, S. A., Rahim, S. A., Ahmi, A., Rahman, N. A. A., & Mazlan, A. U. (2020). Factors influencing sustainable supplier selection: Evidence from palm oil refining and oleochemical manufacturing industry. *International Journal of Supply Chain Management, 9*(1), 437–446.

Vincent, D. (2016). *Privacy: a short history*. John Wiley & Sons: New York.

Visconti, R.M., Larocca, A., and Marconi, M. (2017). Big data-Driven Value Chains and Digital Platforms: from Value Co-Creation to Monetization. *Social Science Research Network*. Available at SSRN: https://ssrn.com/abstract=2903799 or http://dx.doi.org/10.2139/ssrn.2903799.

Wachter, S., Mittelstadt, B. and Floridi, L. (2017). Transparent, explainable, and accountable AI for robotics. *Science robotics*, *2*(6), eaan6080.

Walrave, M., Utz, S., Schouten, A. P., and Heirman, W. (2016). Editorial: The state of online self-disclosure in an era of commodified privacy. *Cyberpsychology, 10*(1), 1. https://doi.org/10.5817/cp2016-1-1.

Wamba, S.F., Akter, S., Edwards, A., Chopin, G. and Gnanzou, D. (2015). How 'big data'can make big impact: Findings from a systematic review and a longitudinal case study. *International journal of production economics*, *165*, pp.234-246.

Wang, C.L., Ahmed, P.K. (2007). Dynamic capabilities: A review and research agenda. *International Journal of Management Reviews*. *9*, 31–51. doi:10.1111/j.1468-2370.2007.00201.x.

Wang, K., Wang, Y. and Yao, J. (2005). A comparative study on marketing mix models for digital products. In *Internet and Network Economics: First International Workshop, WINE 2005, Hong Kong, China, December 15-17, 2005. Proceedings 1* (pp. 660-669). Springer Berlin Heidelberg.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal. 5*: 171-180. https://doi.org/10.1002/smj.4250050207.

Westin, A.F. (1968). Privacy and freedom. *Washington and Lee Law Review*, *25*(1), p.166.

White HD (2004) Citation analysis and discourse analysis revisited. *Applied Linguistics, 25*(1), 89-116.

Wicke, J., & Püster, K. (2019). *Strategische Datennutzung und Datenschutz*. In M. Reich, & C. Zerres, Handbuch Versicherungsmarketing (S. 307-324). Berlin: Springer Verlag GmbH Deutschland.

Wieringa, J., Kannan, P.K., Ma, X., Reutterer, T., Risselada, H. and Skiera, B. (2021). Data analytics in a privacy-concerned world. *Journal of Business Research*, *122*, pp.915-925.

Willing, C., Brandt, T. and Neumann, D. (2017). Electronic mobility market platforms–a review of the current state and applications of business analytics. *Electronic Markets*, *27*(3), pp.267-282.

Winegar, A. G., and Sunstein, C. R. (2019). How much is data privacy worth? A preliminary investigation. *Journal of Consumer Policy, 42*(3), 425–440. https://doi.org/10.1007/s10603-019-09419-y.

Wong, D. (2012). Data is the next frontier, analytics the new tool. *Five Trends in Big Data and Analytics, and Their Implications for Innovation and Organizations*.

Wu, K.W., Huang, S.Y., Yen, D.C. and Popova, I. (2012). The effect of online privacy policy on consumer privacy concern and trust. *Computers in human behavior*, *28*(3), pp.889-897.

Wulf, A.J. and Seizov, O. (2022). "Please understand we cannot provide further information": evaluating content and transparency of GDPR-mandated AI disclosures. *AI & SOCIETY*, pp.1-22.

Wuttke, L. (2020). Datasolut. Von CRISP-DM: *Grundlagen, Ziele und die 6 Phasen des Data Mining Prozess*, Retrived from: https://datasolut.com/crisp-dm-standard/.

Xu, H., Dinev, T., Smith, H., and Hart, P. (2008). Examining the formation of individual's privacy concerns: Toward an integrative view. In *Proceedings of the international conference on information systems* (pp. 6). Paris: ICIS.

Xu, H., Teo, H. H., Tan, B. C. Y., and Agarwal, R. (2009). The role of push-pull technology in privacy calculus: The case of location-based services. *Journal of Management Information Systems, 26*(3), 135–174. https://doi.org/10.2753/mis0742-1222260305.

Yao-Huai, L. (2020). Privacy and data privacy issues in contemporary China. In *The Ethics of Information Technologies* (pp. 189-197). Routledge.

Yoo, Y., Boland Jr, R.J., Lyytinen, K. and Majchrzak, A. (2012). Organizing for innovation in the digitized world. *Organization science*, *23*(5), pp.1398-1408.

Yoo, Y., Lyytinen, K.J., Boland, R.J. and Berente, N. (2010). The next wave of digital innovation: Opportunities and challenges: A report on the research workshop 'Digital Challenges in Innovation Research'.

Young, A.L., and Quan-Haase, A. (2013). Privacy protection strategies on facebook. *Information, Communication and Society, 16*, 479 - 500.

Zhan, Y., Tan, K.H., Li, Y. and Tse, Y.K. (2018). Unlocking the power of big data in new product development. *Annals of Operations Research*, *270*, pp.577-595.

Zhang, J. and Jiang, J.Q. (2001). Sharing information goods and its way of organizing: an economic analysis. *China Economic Quarterly*, *1*(4), pp.937-952.

Zhang, X., Wang, W., de Pablos, P. O., Tang, J., & Yan, X. (2015). Mapping development of social media research through different disciplines: Collaborative learning in

management and computer science. *Computers in Human Behavior, 51*, 1142–1153. https://doi.org/10.1016/j.chb.2015.02.034.

Zhao, F and Collier, A (2016). Digital Entrepreneurship: Research and Practice. The 9th Annual *Conference of the EuroMed Academy of Business*, 2016.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for the future at the new frontier of power. London: Profile Books.

Zuiderveen Borgesius, F. (2015). Improving privacy protection in the area of behavioural targeting. *Available at SSRN 2654213*.

Zupic I., Čater T. (2015). Bibliometric methods in management and organization. *Organizational Research Methods, 18*, 429-472.

German Summary

Die fortschreitende Digitalisierung verändert die Gesellschaft und hat weitreichende Auswirkungen auf Menschen und Unternehmen. Grundlegend für diese Veränderungen sind die neuen technologischen Möglichkeiten, Daten in immer größerem Umfang und für vielfältige neue Zwecke zu verarbeiten. Von besonderer Bedeutung ist dabei die Verfügbarkeit großer und qualitativ hochwertiger Datensätze, insbesondere auf Basis personenbezogener Daten. Sie werden entweder zur Verbesserung der Produktivität, Qualität und Individualität von Produkten und Dienstleistungen oder gar zur Entwicklung neuartiger Dienstleistungen verwendet. Heute wird das Nutzerverhalten, trotz weltweit steigender gesetzlicher Anforderungen an den Schutz personenbezogener Daten, aktiver und umfassender verfolgt als je zuvor. Dies wirft vermehrt ethische, moralische und gesellschaftliche Fragen auf, die nicht zuletzt durch populäre Fälle des Datenmissbrauchs in den Vordergrund der politischen Debatte gerückt sind. Angesichts dieses Diskurses und der gesetzlichen Anforderungen muss heutiges Datenmanagement drei Bedingungen erfüllen: Erstens die Legalität bzw. Gesetzeskonformität der Nutzung, zweitens die ethische Legitimität. Drittens sollte die Datennutzung aus betriebswirtschaftlicher Sicht wertschöpfend sein. Im Rahmen dieser Bedingungen verfolgt die vorliegende kumulative Dissertation vier Forschungsziele mit dem Fokus, ein besseres Verständnis (1) der Herausforderungen bei der Umsetzung von Gesetzen zum Schutz von Privatsphäre, (2) der Faktoren, die die Bereitschaft der Kunden zur Weitergabe persönlicher Daten beeinflussen, (3) der Rolle des Datenschutzes für das digitale Unternehmertum und (4) der interdisziplinären wissenschaftlichen Bedeutung, deren Entwicklung und Zusammenhänge zu erlangen.

Mit Blick auf die Legalität und Legitimität der Datenverarbeitung liefert diese Dissertation Erkenntnisse über die Herausforderungen in der praktischen Umsetzung gesetzlicher Anforderungen, um in der Folge geeignete technische Lösungen abzuleiten. Sie untersucht insbesondere Herausforderungen, die sich aus der Weiterentwicklung von Datenschutzgesetzen ergeben, da die Konformität mit zukünftiger Datenschutzgesetzgebung die Basis datengetriebener Geschäftsmodelle darstellt. Neben Herausforderungen hinsichtlich der Gesetzeskonformität werden Möglichkeiten zur Effizienzsteigerung in den Blick genommen. Die entsprechenden Forschungsarbeiten vermitteln ein besseres Verständnis dafür, wie die Umsetzung der Datenschutzgrundverordnung (DSGVO) verbessert und die kontinuierliche Einhaltung von Datenschutzstandards unterstützt werden kann. So zeigen die ersten drei Forschungsarbeiten, dass eine vollständige Umsetzung der DSGVO derzeit,

insbesondere hinsichtlich der Transparenz- und Dokumentationspflichten, schwer zu erreichen ist. Die Vielzahl und die Dynamik der Verarbeitungsprozesse erschweren eine vollständige und aktuelle Dokumentation. Neue Verarbeitungstechnologien wie Data Mining können eingesetzt werden, um diesen Herausforderungen zu begegnen und Anforderungen gesetzeskonform sowie effizient umzusetzen. Bereits erzielte Ergebnisse in der Organisation können dabei helfen, den Aufwand bei der Implementierung von Data Mining deutlich zu reduzieren.

Mit Blick auf die Wertschöpfung ist ein kontinuierlicher legaler Zugriff auf möglichst viele und hochwertige Daten von großer Bedeutung. In vielen Verarbeitungsprozessen wird nach den Datenschutzgesetzen die stete Kontrolle der Betroffenen über ihre Daten, beispielsweise in Form einer individuellen Einwilligung, gefordert. Aus diesem Grund sind Erkenntnisse über das Nutzerverhalten hinsichtlich der Weitergabe von personenbezogenen Daten erforderlich. Im zweiten Teil der Dissertation werden quantitative Methoden verwendet, um zu einem besseren Verständnis der Einflussfaktoren auf die Bereitschaft von Nutzern, personenbezogene Daten zu teilen, beizutragen. Die Ergebnisse zeigen, dass das Vertrauen gegenüber den Anbietern, das Wissen der betroffenen Personen über die Datenverarbeitung sowie der wahrgenommene Mehrwert des Produkts oder der Dienstleistung wesentliche Einflussfaktoren auf die Bereitschaft sind, Daten zu teilen. Durch gezielte Maßnahmen kann das Management positiven Einfluss nehmen. Dadurch können Unternehmen die Risiko-/ Nutzen-Abwägung der Kunden, um über eine potenzielle Weitergabe ihrer Daten zu entscheiden, zu ihren Gunsten beeinflussen. Das Resultat ist eine höhere Datenverfügbarkeit und somit eine gesteigerte, potenzielle Wertschöpfung.

Im Kontext des dritten Forschungsziels beleuchtet die Dissertation die Rolle des Datenschutzes für digitale Unternehmer und beschreibt die sich aus der Entwicklung der Gesetze ergebenden Geschäftsmöglichkeiten. Datenschutz wird teilweise als Hindernis für wertschöpfende Prozesse oder als signifikantes, rechtliches Risiko wahrgenommen. Die ständige Weiterentwicklung der Gesetze und die Notwendigkeit der Regeleinhaltung bieten jedoch vielfältige Geschäftsmöglichkeiten für Start-ups, die sich auf die Entwicklung von digitalen Datenschutzlösungen und Datenverarbeitungstechnologien konzentrieren.

Ein weiteres Forschungsziel besteht darin, die multidisziplinäre Bedeutung von datenbezogenen Gesetzen am Beispiel der DSGVO besser zu verstehen und Erkenntnisse daraus abzuleiten. Die DSGVO, als Forschungsgegenstand, hat sich innerhalb der Forschungsdisziplinen Recht, Wirtschaft und Informatik unterschiedlich entwickelt. Das Publikationsvolumen von Forschungsartikeln zur DSGVO in den Bereichen Recht, Wirtschaft

und Informatik zeigt dabei ein insgesamt starkes Wachstum im letzten Jahrzehnt, jedoch mit einem zeitlichen Verzug zwischen den Disziplinen. Die Ergebnisse deuten darauf hin (in Bearbeitung), dass durch eine vernetzte, multidisziplinäre Herangehensweise zeitliche Verluste in der Umsetzung reduziert und die Effizienz gesteigert werden können.

Die Dissertation vertieft das notwendige Verständnis für einen kontinuierlichen, rechtskonformen Zugriff auf personenbezogene Daten. Die Ergebnisse sind nicht nur zur Erweiterung des akademischen Diskurses relevant, sondern umfassen gleichermaßen praktische Implikationen für das Management und digitale Start-ups.

Declarations of Co-Authorship

This thesis includes six articles based on jointly authored research. Please find the declarations of my individual research contributions as well as the co-authors' contributions for each article on the following pages. The declarations detail the contributions with respect to both analytical input (e.g., contributions to the conception and design or analysis and interpretation of data) and writing the manuscript (e.g., drafting the article or revising it critically).

Declaration of co-authorship

Publication: Koehler, W., Schultz, C. & Rasche, C. (2020). Das 100% Problem im Datenschutz, 25. G-Forum Jahreskonferenz, Karlsruhe.

This publication was co-authored with Prof. Dr. Christian Schultz and under the supervision of Prof. Dr. Christoph Rasche.

Wolfgang Koehler conceptualized the research project. He wrote all parts of the article and revised the manuscript based on the reviewers' comments.

The contribution of Prof. Dr. Christian Schultz included providing feedback on the initial research idea, the refinement of ideas and editing the manuscript. The contribution of Prof. Dr. Christoph Rasche included providing feedback on the improvement of ideas and revising the manuscript.

I have reviewed the candidate's declaration and the descriptions of his contribution are in accordance with my view of the cooperation.

Declaration of co-authorship

Publication: Guemues, C., Koehler, W., Schultz, C., Rasche, C. (2021). „Konzept eines CRISP-DM-Modells zur ganzheitlichen Datenschutzbetrachtung unter Anwendung von Data Mining". INFORMATIK 2021 - 51st Annual Conference of the German Informatics Society.

This publication was co-authored with Can Gümüs and under the supervision of Prof. Dr. Christian Schultz and Prof. Dr. Christoph Rasche.

Wolfgang Koehler conceptualized the research project. He wrote all parts of the article concerning privacy regulation and privacy processes and revised the manuscript based on the reviewers' comments.

Can Gümüs performed the literature review and wrote all parts concerning the data mining model and its phases.

The contribution of Prof. Dr. Christian Schultz and Prof. Dr. Christoph Rasche included providing feedback on the initial research idea and the refinement of ideas, as well as editing the manuscript.

I have reviewed the candidate's declaration and the descriptions of his contribution are in accordance with my view of the cooperation.

Declaration of co-authorship

Publication: Koehler, W., Schultz, C. & Rasche, C. (2022, forthcoming). The Magic Triangle of Data Governance – Data Governance Insights, Deloitte GmbH (forthcoming 2022)

This publication was co-authored with Prof. Dr. Christian Schultz and under the supervision of Prof. Dr. Christoph Rasche.

Wolfgang Koehler conceptualized the research project. He wrote all parts of the article and revised the manuscript based on the reviewers' comments.

The contribution of Prof. Dr. Christian Schultz included providing feedback on the initial research idea, the refinement of ideas and editing the manuscript. The contribution of Prof. Dr. Christoph Rasche included providing feedback on the improvement of ideas and revising the manuscript.

I have reviewed the candidate's declaration and the descriptions of his contribution are in accordance with my view of the cooperation.

Declaration of co-authorship

Publication: Koehler, W., Schultz, C. & Rasche, C. (2020). "When Handing out Presents is not Enough! – Influencing Factors on the User´s Willingness to Share Data for Connected Car Services", 25. G-Forum Jahreskonferenz, Karlsruhe.

This publication was co-authored with Prof. Dr. Christian Schultz and under the supervision of Prof. Dr. Christoph Rasche.

Wolfgang Koehler conceptualized the research project. He wrote all parts of the article and revised the manuscript based on the reviewers' comments.

The contribution of Prof. Dr. Christian Schultz included providing feedback on the initial research idea, the refinement of ideas, supporting the data analyses and editing the manuscript. The contribution of Prof. Dr. Christoph Rasche included providing feedback on the improvement of ideas and revising the manuscript.

I have reviewed the candidate's declaration and the descriptions of his contribution are in accordance with my view of the cooperation.

Declaration of co-authorship

Publication: Koehler, W., Schultz, C. & Rasche, C. (2022, forthcoming). Legal Access to Data as a Dynamic Capability - The Case of Connected Car Services, Strategic Management Journal (submitted)

This publication was co-authored with Prof. Dr. Christian Schultz and under the supervision of Prof. Dr. Christoph Rasche.

Wolfgang Koehler conceptualized the research project. He wrote all parts of the article and revised the manuscript based on the reviewers' comments.

The contribution of Prof. Dr. Christian Schultz included providing feedback on the initial research idea, the refinement of ideas, supporting the data analyses and editing the manuscript. The contribution of Prof. Dr. Christoph Rasche included providing feedback on the improvement of ideas and revising the manuscript.

I have reviewed the candidate's declaration and the descriptions of his contribution are in accordance with my view of the cooperation.

Declaration of co-authorship

Publication: Koehler, W., Schultz, C., Rasche, C. (2022). „ Data are the Fuel for Digital Entrepreneurship—But what about data privacy?" *The Handbook of Digital Entrepreneurship*, UK: Edward Elgar Publishing Ltd.

This publication was co-authored with Prof. Dr. Christian Schultz and under the supervision of Prof. Dr. Christoph Rasche.

Wolfgang Koehler conceptualized the research project. He wrote all parts of the article and revised the manuscript based on the reviewers' comments.

The contribution of Prof. Dr. Christian Schultz included providing feedback on the initial research idea, the refinement of ideas and editing the manuscript. The contribution of Prof. Dr. Christoph Rasche included providing feedback on the improvement of ideas and revising the manuscript.

I have reviewed the candidate's declaration and the descriptions of his contribution are in accordance with my view of the cooperation.

Statutory Declaration

Eidesstattliche Erklärung

Ich versichere an Eides statt, dass meine hinsichtlich der früheren Teilnahme an Promotionsverfahren gemachten Angaben richtig sind und, dass die eingereichte Arbeit oder wesentliche Teile derselben in keinem anderen Verfahren zur Erlangung eines akademischen Grades vorgelegt worden sind. Ich versichere darüber hinaus, dass bei der Anfertigung der Dissertation die Grundsätze zur Sicherung guter wissenschaftlicher Praxis der DFG eingehalten wurden, die Dissertation selbständig und ohne fremde Hilfe verfasst wurde, andere als die von mir angegebenen Quellen und Hilfsmittel nicht benutzt worden sind und die den benutzten Werken wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht wurden.

Einverständniserklärung

Einer Überprüfung der eingereichten Dissertation bzw. der eingereichten Schriften mittels einer Plagiatsprüfungssoftware stimme ich zu.

Stuttgart den 26.07.2023 (Wolfgang Köhler)

Curriculum Vitae

Diese Seite enthält persönliche Daten und wurde aus der Arbeit entfernt.