



MARTIN-LUTHER-UNIVERSITÄT
HALLE-WITTENBERG



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN IN PUBLICA COMMODA
SEIT 1737



Stephan Kuehnel · Ilja Nastjuk · Stefan Sackmann · Simon Trang (Eds.)

CIISR

Current Information Security and Compliance Issues in Information Systems Research 2023

Proceedings of the 3rd International Workshop on
Current Information Security and Compliance Issues in
Information Systems Research (CIISR 2023)

Co-located with the 18th International Conference on
Wirtschaftsinformatik (WI 2023)

Paderborn, Germany, September 18, 2023

Editors

Dr. Stephan Kuehnel
Martin Luther University Halle-Wittenberg,
06108 Halle (Saale), Germany

Dr. Ilja Nastjuk
University of Goettingen,
37073 Göttingen, Germany

Prof. Dr. Stefan Sackmann
Martin Luther University Halle-Wittenberg,
06108 Halle (Saale), Germany

Prof. Dr. Simon Trang
Paderborn University | University of Goettingen
33098 Paderborn, Germany | 37073 Göttingen, Germany

**Originally published online by CEUR Workshop Proceedings
(CEUR-WS.org, ISSN 1613-0073)**

Originally published in:

*Stephan Kuehnel, Ilja Nastjuk, Stefan Sackmann, Simon Trang (Eds.):
Proceedings of the 3rd International Workshop on Current Information Security and
Compliance Issues in Information Systems Research (CIISR 2023), Co-located with the 18th
International Conference on Wirtschaftsinformatik (WI 2023), Paderborn, Germany, September
18, 2023. CEUR Workshop Proceedings 3512, CEUR-WS.org 2023*

Copyright © 2023 for the individual papers by the papers' authors. Copyright © 2023 for the volume as a collection by its editors. This volume and its papers are published under the Creative Commons License Attribution 4.0 International (CC BY 4.0).

Contents

Prologue

Introduction and Preface to the 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research.....	1
<i>Stephan Kuehnel, Ilja Nastjuk, Stefan Sackmann, and Simon Trang</i>	

Full Papers

Interaction Patterns for Regulatory Compliance in Federated Learning.....	6
<i>Mahdi Sellami, Tomas Bueno Momčilović, Peter Kuhn, and Dian Balta</i>	
A User-centric View on Data Breach Response Expectations.....	19
<i>Felix Hillmann, Tim Klauenberg, Lennart Schroeder, and Till Ole Diesterhöft</i>	
Integrating IT Security Aspects into Business Process Models: A Taxonomy of BPMN Extensions.....	38
<i>Leonard Nake</i>	
From Pixels to Generalization: Ensuring Information Security and Model Performance with Design Principles for Synthetic Image Data in Deep Learning.....	49
<i>Martin Böhmer</i>	

Short Papers

Privacy-Enhancing Technologies in the Process of Data Privacy Compliance: An Educational Perspective.....	62
<i>Alexandra Klymenko, Stephen Meisenbacher, Florian Messmer, and Florian Matthes</i>	
Nudging Towards Compliance? Assessing the Impact of Nudging Strategies on Information Security Policy Adherence.....	70
<i>Theresa Pfaff</i>	
How to Foster Compliance in Non-Integrated IT-Landscapes? The Case of Manual Medical Data Transfers.....	77
<i>Gilbert Georg Hövel and Tizian Matschak</i>	

Extended Abstracts

The Structure of Data Privacy Compliance.....	85
<i>Alexandra Klymenko, Stephen Meisenbacher, and Florian Matthes</i>	

Introduction and Preface to the 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research

Stephan Kuehnel¹, Ilja Nastjuk², Stefan Sackmann¹, and Simon Trang^{2,3}

¹ Chair of Information Systems, esp. Business Information Management, Martin Luther University Halle-Wittenberg, Universitaetsring 3, 06108 Halle (Saale), Germany

² Chair for Information Security and Compliance, Georg August University of Goettingen, Platz der Goettinger Sieben 5, 37073 Goettingen, Germany

³ Chair for Information Systems, esp. Sustainability, Paderborn University, Warburger Straße 100, 33098 Paderborn, Germany

Abstract

This volume contains the proceedings of the 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research (CIISR 2023), held at the 18th International Conference on Wirtschaftsinformatik (WI 2023) in Paderborn, Germany, on September 18, 2023.

Keywords

CIISR 2023, WI 2023, Information Security, Compliance, IT, ISR

1. Introduction


In a connected world of people, data, and things, enterprises are caught between the need for rapid digital growth, regulatory compliance, and securing their information assets across all stakeholders [1]. Effective compliance and security governance as well as the appropriate implementation of corresponding measures are becoming a central factor for digital responsibility and sustainable security [2].

Nowadays, information security and compliance are approached from a variety of different perspectives in information systems research (ISR). As part of information security management, for instance, it is examined which operational measures may result in desired employee behavior [1, 3]. In the context of cloud computing, for instance, it is examined how compliance with service-level agreements can be achieved in hybrid cloud architectures [4]. In the context of business process management, for instance, it is examined how information security and compliance measures in business processes can be ensured sustainably and economically in digitalized and electronic markets [5, 6].

As part of the third edition of this workshop, we acknowledged the thematic link between compliance and information security and decided also to reflect this in the title of the workshop, which is now called the *International Workshop on Current Information Security and Compliance Issues in Information Systems Research (CIISR)*. This year's edition, held on September 18, 2023, in conjunction with the 18th International Conference on Wirtschaftsinformatik in Paderborn, Germany, consisted of several presentations and a poster session. Based on the main theme of the conference—DIGITAL RESPONSIBILITY—we discussed current issues related to the


CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ stephan.kuehnel@wiwi.uni-halle.de (S. Kuehnel); ilja.nastjuk@wiwi.uni-goettingen.de (I. Nastjuk); stefan.sackmann@wiwi.uni-halle.de (S. Sackmann); simon.trang@uni-paderborn.de (S. Trang)

 0000-0002-6959-9555 (S. Kuehnel)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

responsible handling of information security and compliance, which are of great importance for ISR in an increasingly digitalized world.

2. Target Group, Submission Types, and Paper Selection

The target audience of the CIISR workshop are scientists whose research focuses on current information security and compliance issues, practitioners working in the field of information security and/or compliance, and all other interested parties. This workshop provides the opportunity for (senior) researchers and practitioners to present their latest findings but also serves as a forum for young scientists and doctoral students to present early or ongoing research results.

We invited authors to submit empirical studies, systematic literature reviews, design science research papers, as well as practitioner papers related to the workshop theme, e.g., information security and compliance at the interface with business processes, cloud computing, or current events such as the COVID-19 pandemic, as well as current challenges in the context of IT compliance and information security policies. We called for submissions from the subject areas listed above that fell into one of the following three submission categories:

1. Full papers (research papers/practical reports)

This submission type includes both advanced research with at least partial evaluation and comprehensive practical contributions.

2. Short papers (research in progress papers/short practical reports)

Short papers represent ongoing research or ongoing practical projects. In addition to presenting initial results, these papers should also contain an outlook on further research or further project progress, including planned future work steps.

3. Extended abstracts

Extended abstracts present and discuss high-quality results of already published contributions (or dissertations/postdoctoral theses) with relevance to the workshop topic.

Full papers were not allowed to exceed 12 pages in the submitted version, and short papers as well as extended abstracts were not allowed to exceed six pages, including title, abstract, and placeholders for author information and acknowledgments. The bibliography and appendices were not included in the page count.

Full and short papers were subjected to rigorous double-blind review by two reviewers, where at least one of the reviewers was a member of the Program Committee. Extended abstracts were reviewed single-blind. All reviews focused on five criteria: 1) quality of the theoretical contribution, 2) appropriate use of research methods, 3) degree of innovation and significance of the contribution, 4) presentation and language, and 5) potential of the contribution to foster discussion. Program Committee members were asked to make recommendations to accept, revise, or reject the submissions, which were then discussed by the four workshop chairs to arrive at the final decisions.

A total of 11 papers were submitted for the workshop, of which one full paper was directly accepted and seven were accepted under conditions. Authors of full papers were allowed an additional two pages to incorporate reviewer comments, and authors of short papers and extended abstracts were allowed one page each. In addition to the final version of each paper, a response letter was required to provide information on how and to what extent the reviewer comments were addressed. After another review of the papers and the response letters by the workshop chairs, four full papers, three short papers, and one extended abstract could be accepted. The acceptance rate for full papers was 80% and 60% for short papers. In addition, the extended abstract was also accepted.

3. Contents of the CIISR 2023 Workshop

In line with WI 2023, the CIISR workshop was held locally in Paderborn, Germany. In total, more than 30 participants have registered. The CIISR 2023 workshop and these workshop proceedings include 8 papers:

1. The full paper **Interaction Patterns for Regulatory Compliance in Federated Learning** written by Mahdi Sellami, Tomas Bueno Momčilović, Peter Kuhn, and Dian Balta deals with federated learning (FL), where organizations share local machine learning models while the data remain on-premise. For this context, the paper develops four interaction patterns that enable compliance-by-design and trust-context-sensitive analyses of an FL system by combining different privacy-preserving approaches.
2. The full paper **A User-centric View on Data Breach Response Expectations** by Felix Hillmann, Tim Klauenberg, Lennart Schroeder, and Till Ole Diesterhöft focuses on individual customer expectations after data breaches in different situations and business environments. Building on prior research on data breaches that have been integrated into expectation confirmation theory, individual customer expectations are analyzed by conducting twelve qualitative interviews. The findings reveal the individual nature of customer expectations about data breach responses, which are shaped by multiple factors.
3. The full paper **Integrating IT Security Aspects into Business Process Models: A Taxonomy of BPMN Extensions** written by Leonard Nake deals with Business Process Model and Notation (BPNM) extensions from the information/IT security domain. Based on a systematic literature review, a taxonomy is developed that provides an overview of common features and dimensions of security-related BPMN extensions and provides profound insights into existing work.
4. The full paper **From Pixels to Generalization: Ensuring Information Security and Model Performance with Design Principles for Synthetic Image Data in Deep Learning** authored by Martin Böhmer deals with the effective and ethical use of synthetic image data for deep learning in computer vision. Based on challenges in obtaining real training data, design principles for the selection, generation, and integration of synthetic images are proposed, including aspects such as ethical compliance, privacy protection, scene diversity, and complexity management.
5. The short paper **Privacy-Enhancing Technologies in the Process of Data Privacy Compliance: An Educational Perspective** by Alexandra Klymenko, Stephen Meisenbacher, Florian Messmer, and Florian Matthes explores the educational needs of practitioners working in the field of data privacy compliance. Drawing on 11 semi-structured interviews and a survey of 24 respondents, the study discusses the learning goals of privacy-enhancing technologies and explores how these goals can be aligned with practitioners' role-specific needs.
6. The short paper **Nudging Towards Compliance? Assessing the Impact of Nudging Strategies on Information Security Policy Adherence** by Theresa Pfaff explores how employee behavior towards information security policy compliance can be influenced by the concept of nudging. The core of the paper is the presentation of a research model that will be used in future research to investigate the effectiveness of nudging strategies as part of an online experiment.

7. The short paper entitled **How to Foster Compliance in Non-Integrated IT-Landscapes? The Case of Manual Medical Data Transfers** written by Gilbert Georg Hövel and Tizian Matschak addresses the issue that medical professionals often have to manually transfer medication data between different health information systems, which can lead to errors with serious consequences for patients (medication non-compliance). The paper presents a research design that will be used in future research to investigate how different formal sanction mechanisms of deterrence theory relate to different types of medication errors.
8. The extended abstract **The Structure of Data Privacy Compliance** by Alexandra Klymenko, Stephen Meisenbacher, and Florian Matthes deals with data privacy compliance and interprets it as a dynamic process that depends on the roles involved and the nature of their interactions. Based on the results of a previously published interview study, the extended abstract briefly presents a graphical structure that maps the various roles and interactions diagrammatically.

4. Organization and Acknowledgement

The workshop organization lay in the hands of Dr. Stephan Kuehnel (workshop chair and web chair), Dr. Ilja Nastjuk, Prof. Dr. Stefan Sackmann, and Prof. Dr. Simon Trang (workshop co-chairs). We would like to express our deepest gratitude to the members of the Program Committee for their active participation in the review and paper selection process:

- Prof. Dr. Jörn Altmann (Seoul National University, South Korea)
- Prof. Dr. Alfred Benedikt Brendel (TU Dresden, Germany)
- Prof. Dr. Nadine Guhr (OWL University of Applied Sciences and Arts, Germany)
- Ass. Prof. Dr. Simon Hacks (Stockholm University, Sweden)
- Dr. Kristin Masuch (University of Göttingen, Germany)
- Mohammed Mubarkoot, Ph.D. (Seoul National University, South Korea)
- Prof. Dr. Jana Rhese (University of Mannheim, Germany)
- Prof. Dr. Michael Schulz (NORDAKADEMIE Hochschule der Wirtschaft, Germany)
- Michael Seifert, M.Sc. (GISA GmbH, Germany)
- Dr. Tobias Seyffarth (Federal Office for Information Security, Germany)
- Prof. Dr. Nils Urbach (Frankfurt University of Applied Sciences, Germany)

We would also like to thank the additional reviewers and sub-reviewers Laura Bauer, Martin Böhmer, Johannes Damarowsky, Gilbert Georg Hövel, Julia Klein, Luis Laemmermann, Tizian Matschak, Leonard Nake, Theresa Pfaff, and Florian Rampold for their active support as well as the organizers and the staff of the 18th International Conference on Wirtschaftsinformatik for including our CIISR Workshop in the conference program and for their continued assistance in organizational and technical matters. Last but not least, we are grateful to all the speakers, poster presenters, and participants who made the CIISR Workshop 2023 a great event.

References

- [1] S. Trang, B. Brendel, “A Meta-Analysis of Deterrence Theory in Information Security Policy Compliance Research,” *Information Systems Frontiers*, vol. 21, no. 6, pp. 1265–1284, 2019.
- [2] D. Schatz, R. Bashroush, “Economic valuation for information security investment: a systematic literature review,” *Information Systems Frontiers*, vol. 19, no. 5, pp. 1205–1228, 2017.

- [3] S. Hengstler, S. Kuehnel, K. Masuch, I. Nastjuk, S. Trang, "Should i really do that? Using quantile regression to examine the impact of sanctions on information security policy compliance behavior," *Computers & Security*, vol. 133, p. 103370, 2023.
- [4] M. Seifert, S. Kuehnel, S. Sackmann, "Hybrid Clouds Arising from Software as a Service Adoption: Challenges, Solutions, and Future Research Directions," *ACM Computing Surveys*, vol. 55, no. 11, pp. 1–35, 2023.
- [5] T. Seyffarth, S. Kuehnel, "Maintaining business process compliance despite changes: a decision support approach based on process adaptations," *Journal of Decision Systems*, vol. 31, no. 3, pp. 305–335, 2022.
- [6] S. Sackmann, S. Kuehnel, T. Seyffarth, "Using Business Process Compliance Approaches for Compliance Management with Regard to Digitization: Evidence from a Systematic Literature Review," in *Business process management: 16th International Conference, BPM 2018, Sydney, NSW, Australia, September 9-14, 2018: proceedings*, M. Weske, M. Montali, I. M. Weber et al., Eds., vol. 11080, pp. 409–425, Springer, Cham, 2018.

Interaction Patterns for Regulatory Compliance in Federated Learning

Mahdi Sellami¹, Tomas Bueno Momčilović¹, Peter Kuhn¹ and Dian Balta¹

¹ fortiss GmbH, Guerickestraße 25, Munich, Germany

Abstract

Organizations in highly regulated domains often struggle to build well-performing machine learning (ML) models due to restrictions from data protection regulation. Federated learning (FL) has recently been introduced as a potential remedy, whereby organizations share local models while keeping data on premise. Still, regulatory compliance remains challenging in FL settings: training data needs to be shared to some extent, and models can be reverse engineered or misused towards violation of data privacy by each participating organization. Guided by design science methodology, we introduce four interaction patterns that allow for compliance-by-design and trust-context-sensitive analysis of an FL system by combining different approaches to privacy preservation. We match the patterns to privacy principles and exemplify how verifiable claims about compliance at design- and operation-time FL can be generated to make all participating organizations accountable.

Keywords

Federated Learning, Privacy, Compliance, Design Patterns

1. Introduction

Organizations in highly regulated domains, such as the government, health or banking, explore applications of machine learning (ML) with high intensity to gain efficiency, effectiveness, and a competitive edge (cf. e.g., [1-2]). Unfortunately, they often struggle with having a sufficient amount of data [3]. This insufficiency of data is often the leading cause of underperforming models of ML [4], thereby undermining the value proposition.

One remedy is to federate the learning process between different organizations and their data, in order to train a common (and higher quality) model that benefits the knowledge of all parties involved (cf. e.g., [4]). A prerequisite for such an approach is that the parties have to share some of the data or at least model training parameters. This is not an easy task, given strong regulatory constraints resulting from e.g., GDPR, HIPAA, and similar (cf. e.g., [5]). Implementing federated learning (FL) would mean that they have to be compliant, i.e. that 1) they have a corresponding, suitable design that complies with regulation; and 2) they operate according to it. While compliance needs to be upheld throughout the whole process of setting up the federation, model training and processing, an approach is required to design corresponding information systems (IS) that hold every participating organization accountable to preservation of privacy in every stage of ML.

In this paper, we address the following question: *How to design an architecture for accountable privacy-preserving data sharing in the entire ML process?* We propose **interaction patterns** for the architecture design of IS involved in inter-organizational ML with a focus on the level of trust between organizations. In terms of implications, the patterns lay the ground for (1) an academic discussion in mapping legal regulation requirements to technology, and (2) a practical guideline

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ sellami@fortiss.org (M. Sellami); momcilovic@fortiss.org (T. Bueno Momčilović); kuhn@fortiss.org (P. Kuhn); balta@fortiss.org (D. Balta)

ORCID 0000-0002-2817-2643 (M. Sellami); 0000-0003-4503-2244 (T. Bueno Momčilović); 0000-0001-6774-2904 (P. Kuhn); 0000-0001-8311-3227 (D. Balta)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

for designing such systems. In order to evaluate the patterns, we present a prototypical implementation and discuss how one can make and prove claims of what was designed and promised.

2. Background

2.1. Federated Learning

Federated learning (FL) is a novel machine learning (ML) method that allows a set of distributed parties to jointly train a shared model while keeping their data on premise. The FL process involves an aggregator who provides specifications for applying an ML model, sends them to the parties to train the model on their own private data, and then combines the information from many different local models using an algorithm for aggregating training parameters [4]. The most common approach relies on a client-server or star network [6] comprising the Federated Averaging algorithm proposed by [4], where a central server orchestrates the contributions of clients (participants such as organizations or edge devices).

FL system designs vary in six main aspects [7]: 1. data partitioning (horizontal vs. vertical), 2. ML model, 3. scale of federation (cross-silo vs. cross-device), 4. communication architecture (centralized vs. decentralized), 5. privacy mechanism, and 6. motivation. To illustrate our results, we limit our scope to: 1. horizontal data partitioning; 2. neural network; 3. cross-silo scale; 4. centralized communication; 5. no privacy mechanism beyond the learning pipeline; and 6. a motivation for federation primarily driven by the need to comply with data protection regulation. Furthermore, we consider that the FL process contains four stages [8]: **setup** to specify the data, purpose of collaboration, and the ML pipeline; **preprocessing** to prepare the training data of each participant; **processing**, to optimize the model iteratively; and optional **postprocessing** for applying a privacy-enhancing technique, mitigate bias and similar.

With regard to compliance, **privacy mechanisms** are of particular importance. Even though each party's data remains on-premise during training, attackers may still extract sensitive information from the exchanged model updates. Several attacks including membership inference attacks [9] and model inversion attacks [10] can lead to data leakage. Thus, different privacy techniques (e.g., differential privacy [11]; cryptographic methods [12]) and trusted execution environments [13] can be independently combined with FL to provide stronger privacy guarantees. The **motivation** for FL is also an important compliance requirement in real-world scenarios, ensuring the involvement of participating parties. This motivation is based on regulations, incentives, or a combination of both. Applying FL within an organization (e.g., governments, companies, etc.) is generally driven by a need to comply with regulation [5].

2.2. Compliance based on accountability and verifiable claims

Interpreting and complying with legal requirements is a resource-intensive problem [14]. Accountability helps to address this challenge in FL setups. Implying different domains and stakeholders [15], accountability has been introduced in the 1960s as an important design principle for systems [16], with different operationalizations emerging throughout the decades since (e.g., "code is law"; [17-18]).

In this work, we consider accountability as a mechanism for enabling trust and legal compliance in FL systems. We adopt a definition that is applicable to distributed systems [19], i.e. *"the transparent assignment and ownership of responsibilities (...) enabling [the distribution of] business goals across multiple organizations."* We tackle accountability from an engineering perspective by introducing verifiable claims to FL systems, aiming to-wards trustworthy AI development [20] by emphasizing the following dimensions [8]:

- **Verifiable:** Every step of the learning process must be documented by specific claims. Each claim should be transparent and supported by evidence, that the corresponding step was conducted correctly with respect to a predefined specification.

- **Undeniable consent:** The executed learning process must be aligned with the expectations of all participants, who must give their explicit consent to the specification of the process. Furthermore, the execution of the process should be non-repudiable and provable.
- **Auditable:** Any deviation from the specification (e.g., system attacks) must be detectable and provable by any third party, based on the recorded claims and their corresponding evidence.
- **Tamper-evident:** All the interactions between the participants must have a corresponding record according to a predefined specification. Furthermore, any intended corruption of the shared knowledge of the participants should be detectable.

2.3. Data Protection and Privacy Principles

Privacy-related regulatory requirements in the EU center almost exclusively on the General Data Protection Regulation (GDPR; [21]), which came into force in 2016 and repealed an earlier framework from 1995. The draft EU Act on Artificial Intelligence [22] directly refers to the GDPR for all privacy-related requirements in AI systems development and deployment (Sec. 3 Sub. 2 Art. 10 Para. 5), and so do national laws (e.g., German Federal Data Protection Act; [23]) that harmonize (i.e., adapt) the stipulations to the national context.

GDPR is concerned with the protection of personal data, i.e., “any information relating to an identified or identifiable natural person” (Art. 4 Paragraph 1). It comprises six core **privacy principles** in Article 5. **Lawfulness, fairness, and transparency** refers to obtaining consent from the data subject and defining the legitimate reason for processing personal data. **Purpose limitation** refers to specifying explicit and unexceedable boundaries for processing data, whereas **data minimization** refers to explicit and unexceedable boundaries for collecting data. **Accuracy** refers to keeping data up-to-date and rectifying deviations, and **storage limitation** refers to specifying an explicit time limit when storing collected data. **Integrity and confidentiality** represent security “using appropriate technical or organizational measures.” Finally, the point on **accountability** designates the data controller as responsible for ensuring compliance with the six principles.

These principles in GDPR correspond to a widely accepted set of best privacy practices. Legal scholars [24] provide seven overarching principles, of which six directly map to GDPR: respect for context with purpose limitation; consent, legitimacy and transparency with lawfulness, fairness and transparency; transparency with accuracy; proportionality with data minimization; and accountability with its GDPR counterpart. The unique principle that remains is **privacy by design** – a requirement to address data privacy concerns in the initial design stages and throughout the whole lifecycle of products, processes, and services, which is compatible with the notion of data protection by design in GDPR (Art. 25 paragraph 1, GDPR). Thus, a set of eight principles comprises the privacy requirements from the regulatory standpoint.

3. Research Approach: Pattern-based Design Research

We follow the pattern-based design research (PDR) method [25] to specify reusable patterns for privacy-preserving interactions between at least two parties who want to share knowledge, but not their private data. Patterns are “best practices that are bound to a specific context in which the provided solution has been proven to work” (p. 75, [25]); they represent empirically founded models that are defined in iterations between theory and practice.

Interaction patterns is a term we propose to describe an approach in information systems research for modelling a controlled sequence of information flow between parties with predefined roles (see, e.g., [26-27]). **Interaction** refers to an exchange of information between two or more parties for the purpose of achieving a common goal, whereas **interaction patterns** are design patterns which describe recurring interactions. These interactions correspond to the following simplified scenario of interest. The data processor wants to use the private data of the

data provider for some computation: e.g., data aggregation or pattern recognition. The data provider wants to make sure its private data is protected and correctly handled (e.g., that there is no data leakage).

The process of PDR comprises four stages [25]. First, input is collected from an existing scientific foundation or practical observations. Second, this input is used to generate pattern candidates, their description language, and/or the design theories to support the design activities. Third, pattern candidates are instantiated as tangible solutions to practical problems, and these instances may deviate from each other based on the context they are applied in. Finally, these deviations are evaluated and used as further input to refine existing candidates or define new patterns for the next cycle.

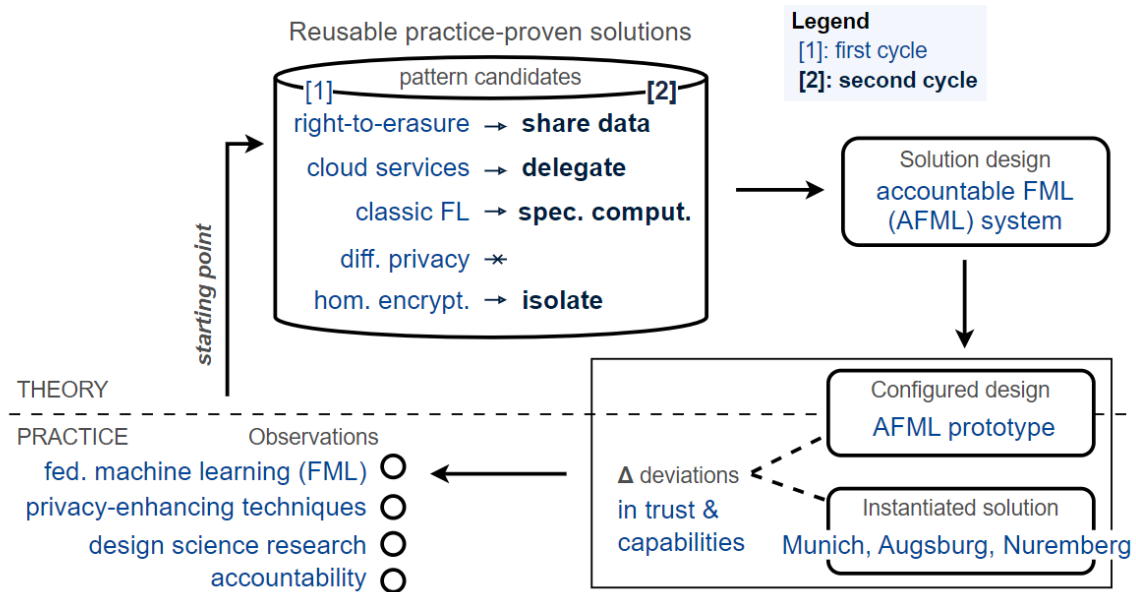


Figure 1. Annotated PDR diagram; adapted from p. 78, Buckl et al. (2013)

We completed an entire cycle of the PDR method, finishing with new pattern candidates (see: Figure 1). We compiled relevant ‘observations’ from a literature review: FL and classic ML methods [3]; privacy enhancing techniques for ML tasks [5]; design science research, a discipline for studying software architectures [25]; and accountability through verifiable claims [28]. With this input, we specified five initial pattern candidates, which reflect the general solutions that are commonly applied to privacy-oriented problems (e.g., [29]).

1. **Right to erasure:** One party shares its data with another party, under the expectation that the latter will delete the data once the purpose for which it was shared no longer exists (Article 17 paragraph 1, [21]).
2. **(Trusted) Cloud services:** A trusted entity hosts the data, performs the computation for all parties, and then deletes the data [30].
3. **Classic FL:** One party sends the model specifications to another party in the processing stage, and aggregates the results from training [4].
4. **Differential privacy:** Parties hide personal identifiable information with type-appropriate noise that does not affect ML model results [31].
5. **Homomorphic encryption:** Parties hide personal identifiable information by transforming it into ciphertext without affecting ML model results [32].

Using privacy by design (i.e., privacy in all stages; [24]) and accountable FL (i.e., federation that can be verified; IBM, [8]) as general guidelines, we created a blueprint of an “accountable federated machine learning” (AFML) system. With it, we developed a prototype for conducting experiments with three Bavarian municipalities – Munich, Augsburg, and Nuremberg.

Input from the experiments and stakeholders helped us restructure the pattern candidates into four patterns which we present below. First, we excluded differential privacy once we

determined that adding noise does not solve the trust problems between the parties themselves, nor excludes the need for data erasure. Second, we added the precondition of secure communication, as successful man-in-the-middle attacks invalidate any privacy claim. Third, we abstracted away from specific implementations. Cloud services are an instance of a pattern of delegating to a third party, and so is homomorphic encryption an instance of a pattern of isolating information without affecting the results. The classical FL has been extended to also include pre- and post-processing specification of computation. Right-to-erasure has become a data sharing pattern, since according to [21], data providers can share data with other parties (e.g., processors), but must delete and enforce deletion of all relevant instances upon a user’s request. In ML scenarios, this can involve running a resource-intensive process to “unlearn” (i.e., retrain) any model to exclude the user’s data [41].

We added four dimensions to distinguish patterns: the trust scenario, architectural model, claim, and exemplary application. **Trust scenarios** (similar to threat models; cf. e.g., [5]) describe the trust between the parties, and the challenge that the pattern is addressing. The **architectural model** is a diagram visualized in design science templates (4+1 view; UML), and contains the systems, the actors, and their interactions. **Claims** are short texts describing the promised way of private data handling, which can be linked with verifiable evidence. Finally, the **exemplary application** provides the instance of the pattern in industry or research.

4. Interaction Patterns

4.1. Patterns

In this section, we introduce four interaction patterns as solutions for private data handling issues: share data, specify computation, delegate and isolate. The visual structure in Figure 2 represents the minimal requirements for the interaction to be completed correctly, such as: necessary actors (i.e., the data provider, the data processor and in one case, a third party), technical components (i.e., modules for computation and a database of private data), and sequential actions for the exchange and processing of data. The names of the patterns reflect the overarching process taking place. Additionally, when implemented correctly, each pattern generates a privacy claim against which we verify that the private data indeed remains private, while maintaining the important assumption that parties have established secure communication.

The *share data* pattern relies on the provider trusting the processor to handle its data properly (i.e., behaves in a trustworthy manner). This pattern is widely observed in highly regulated areas like healthcare and government. In fact, since the introduction of the “right to erasure” or the “right to be forgotten” in the GDPR, this pattern has been adopted by a variety of developers whose applications depend on private data, or it is at least offered to the users as an option.

The *specify computation* pattern also relies on trust. Here, the processor trusts the provider (and its computation system) to execute the computation correctly: with the right data, in the prearranged manner, and without tampering of results. This is exemplified in a classical FL scenario. In the *delegate* pattern, by contrast, neither the provider nor the processor trust each other, but they both trust a third party. This party can be any entity which is deemed trustworthy enough to store and handle data securely, and execute computation steps. Put simply, the core parties shift their trust to an intermediary.

Finally, the *isolate* pattern requires the parties to trust the technology instead of one another. Although this can solve concerns related to trust, it is the most complex and resource-intensive approach. For example, homomorphic encryption [32] allows the provider to encrypt its data before sending it to the processor, who then performs the computation on ciphertext (i.e., the encryption space). The resulting output, when decrypted, equals the output of the computation in plaintext (i.e., original space). However, fully homomorphic encryption methods are still not practical due to storage, configuration, and efficiency issues [33].

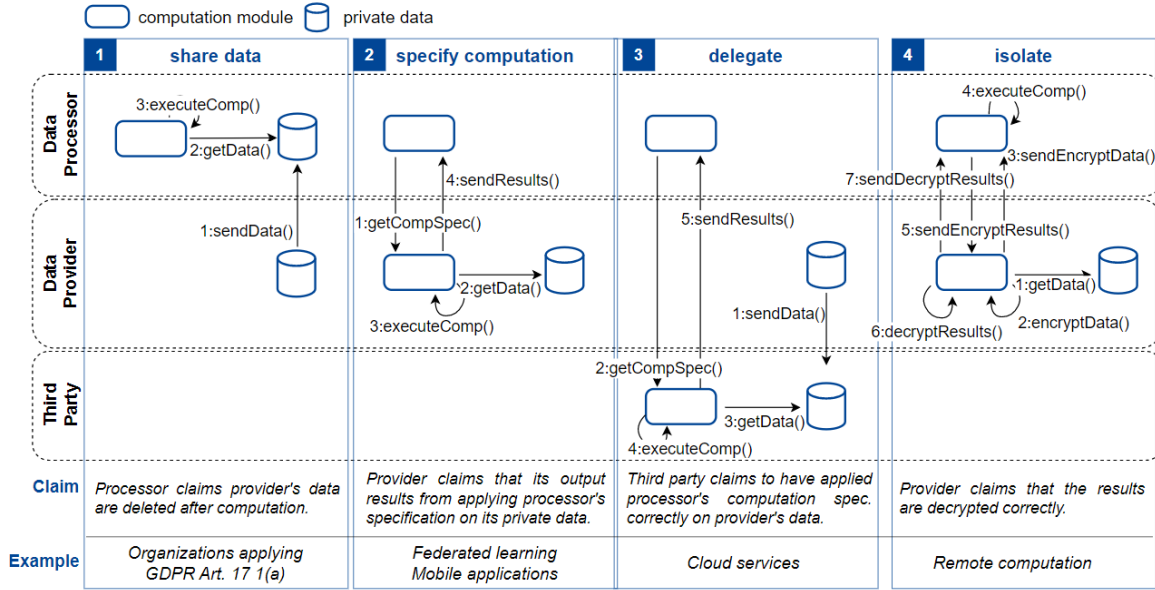


Figure 2. Interaction patterns with corresponding architecture, roles, claims and examples

4.2. Deciding Criteria and Preconditions

The additional outcome of refining pattern candidates has been a small framework for decision-making. Namely, deciding on applicable patterns depends on three criteria and five preconditions. Regarding criterions, first, the level of **trust** between parties is key. If the provider and processor trust that the other has capability and intent to handle information accordingly, then sharing the data or the model parameters is always a possibility. In other words, a simple verifiable promise that the data will be deleted or the computation will be executed suffices. Second, if either one or both parties are mistrustful, information sharing can still be intermediated by a **third party**. Finally, if no third party is trusted or available, more privacy-protective ‘trustless’ solutions are needed (cf. e.g., [29]). However, as Table 1 shows, introducing **complexity** (such as homomorphic encryption) is only justifiable in the strictest cases, because multiple options are available in less strict contexts [5].

Table 1

Trust scenarios (A-I) with available interaction patterns (1-4), sorted by complexity

Trust Scenarios	Processor trusts party X with computation		
Provider trusts party X with data	Provider trusted	Provider not trusted, 3 rd party trusted	Neither party trusted
Processor trusted	A: 1,2,3,4	B: 1,3,4	C: 1,4
Processor not trusted, 3 rd party trusted	D: 2,3,4	E: 3,4	F: 4
Neither party trusted	G: 2,4	H: 4	I: 4

Regarding preconditions, first, neither party has both the necessary components (i.e., sufficient data and computation specifications) nor the capability (i.e., data collection or processing capacity) to execute the process; thus, they have complementary roles [4]. Second, vulnerabilities are substantial enough (e.g., personally identifiable information cannot be anonymized without information loss) to prevent the parties from applying an easy solution. Third, the expected value is high enough to incentivize parties to collaborate (e.g., generated output is significantly more useful than the raw data itself; e.g., [1]). Fourth, expected costs of non-compliance are high enough to disincentivize it (e.g., punishment under GDPR that can reach up to 4% of revenue). Fifth and final, secure communication prevents man-in-the-middle attacks but is itself insufficient for verifying privacy is protected.

5. Exemplary Application of the Interaction Patterns

We evaluated the applicability of the interaction patterns during an FL project for German public services. In collaboration with Munich, Augsburg and Nuremberg, we designed and configured a system based on IBM FL [34] and used it to train a neural network for a multi-class text classification task. The dataset encompasses textual user feedback on their experiences and suggestions of using the German online public services as well as the usability rating in an ordinal range from 0 to 5. In our work, we used the raw text as input for the model and the categories as output (i.e., prediction). The municipalities are interested in collaborative training to automatically forward future feedback to the department corresponding to the category, such that, e.g., feedback about the user interface of the service is forwarded to the user experience (UX) team. Since the free texts may contain sensitive information, and it is not possible to detect all identifier entities (see, e.g.: [35]), we used the interaction patterns to provide data privacy guarantees.

The setup stage included three workshops with the cities to define the learning task, select the learning features of the data, and choose the architecture of the ML model. It is the only stage where an exchange of sensitive data is not needed. We used a character-level convolutional neural network as model architecture, inspired by the work of [36] and conducted two main training experiments. The first experiment was conducted on one municipality dataset to prove the usability of ML and the model architecture, with the accuracy of the fine-tuned model reaching 77,8%. The second experiment involved the FL system to train the model with all three datasets in a federated manner. The model's accuracy improved to 93,8%, confirming the hypothesis that sharing the data of the cities enhances the performance of the model.

Table 2
Application of the patterns 1 and 2 in the pre-processing and training stages

Dimensions		Stages	
		Preprocessing	Processing (training)
Pattern		1. Share data	2. Specify computation
Trust Scenario		[C:1,4] Municipalities do not have the ML proficiency to execute the preprocessing stage by themselves. By signing a data protection agreement and assuring them of access and usage (i.e., purpose) controls, we (data processors) acquired the trust.	[G:2,4] Municipalities involved the IT departments to integrate the system, ensuring the necessary infrastructure for local training is provided. We provided containerized applications as a form of specification. That has been enough to trust them to perform computation correctly.
Roles	Processor	fortiss: We provided expertise to preprocess the data as specified in the setup.	fortiss: We provided expertise to set up and orchestrate the FL system.
	Provider	Municipalities: They hold the training data and provide it as input for the learning task.	Municipalities: They provide the data and IT expertise to train the model locally, and send model weights to us.
	3 rd Party	n/a: Trust is enabled by the agreement.	n/a: Trust is enabled by containerization.
Evidence for Claims (Accountability)		Documenting artifacts (e.g., intermediate results and metadata) and events (e.g., data deletion); in a Factsheet [40].	Documenting artifacts (e.g., model weights and evaluation metrics) and events (e.g., model convergence); in a Factsheet [40].

We applied the patterns in the preprocessing and training stages (cf. Table 2), as these stages involved sensitive data. Postprocessing was not included, as the project did not require any additional (e.g., robustness or fairness) checks. For preprocessing, we used the *share data* pattern because at the time, the municipalities lacked the necessary expertise or resources to preprocess their data. With the help of a contract and the verifiable claim that we will delete the data after the goal is completed, we received the raw data from the municipalities as an upload to a secure cloud. Preprocessing involved excluding the empty rows, cleaning the freetext from special characters, and fixing the misspelled words, with the help of the *pandas* Python package.

For the training, we applied the *specify computation* pattern. Taking the role of the aggregator, we specified computations by providing a containerized application using Docker for three municipalities to execute. The training of local models has been performed using the *keras* interface of the *tensorflow* Python package and aggregated into a common model according to steps in [34], using. Given the incomplete IT infrastructure at the time, we simulated the scenario of FL: we ran the application containers and provided the preprocessed training data as input to assure that the training is federated. In future training sessions, cities will have configured the IT infrastructure.

6. Discussion

6.1. Evaluation

Interaction patterns ideally help organizations pursue privacy-compliant ML. However, knowing which interaction to set up is not enough to satisfy the proposed principles (cf. Section 2.3). The table below evaluates whether a pattern satisfies a principle automatically or needs an additional mechanism to do so. Such mechanisms include internal policies within an organization with penalties for non-compliance; legally enforceable contracts for external entities; or other unspecified mechanisms. Since data providers are ultimately accountable, the onus is on them to set and enforce mechanisms.

In the **principle of lawful processing**, if consent for data collection from users is not obtained (Article 18 paragraph 1 point (a), [21]), or the reason for processing is not considered legitimate, patterns cannot help. In our case, legitimacy lies in “*the performance of a task carried out in the public interest*” (Article 18 paragraph 1 point I, [21]), which was already present before our involvement. The legality of the ML application is an implicit precondition for which provider is ultimately accountable.

Purpose and data limitation show that limits are easier to enforce when data remains on-site. *Specify computation* pattern requires a policy to ensure that the dataset is filtered beforehand, and that the computation parameters do not exceed the predefined purpose. When data is forwarded off-site, the provider can ensure that it has been filtered, but otherwise only enforce compliance with a contract. By contrast, the *isolate* pattern (when instantiated as homomorphic encryption) automatically ensures that an honest-but-curious, infected, or malicious processor cannot see the output, infer what the input consists of, or stretch the processing beyond the predefined purpose.

Table 3
Additional mechanisms each pattern needs for each principle in one stage

Principles	Patterns			
	share data	specify computation	delegate	isolate
Lawful processing	<i>not enough</i>	<i>not enough</i>	<i>not enough</i>	<i>not enough</i>
Purpose limitation	contract	policy	contract	<i>automatic</i>
Data minimization	policy	policy	policy	<i>automatic</i>
Accuracy	contract	policy	contract	policy
Storage limitation	contract	policy	contract	policy
Integrity & confidentiality	<i>not enough</i>	<i>not enough</i>	<i>not enough</i>	<i>automatic</i>
Accountability	contract	policy	contract	policy
Privacy by design	contract	<i>automatic</i>	contract	<i>automatic</i>

Accuracy and storage limitation deal with data maintenance, which are not directly addressed by interaction patterns. Instead, the data provider must have policies for repairing inconsistencies and deleting data after the allotted time. Integrity and confidentiality require more intensive data protection, which is only partly achieved by secure communication. Retaining the data on-site lowers the security risk from external attacks, but the provider can still be vulnerable to reconstruction attacks from an honest-but-curious or malicious processor – i.e., reconstructions of raw data points from model parameters or aggregate information [31]. As the processor is also at risk of being infected or having its results exploited, a risk assessment would be needed to determine vulnerability to such scenarios, and the selection of mechanisms that would guarantee privacy alongside the appropriate interaction pattern (e.g., differential privacy; [31]). If the collected data is immediately *isolated*, however, the only security measure that is needed is the protection of private keys.

Finally, privacy by design can be interpreted in two ways. Using a relaxed interpretation, a provider who demonstrates a conscious and institutionalized concern for protecting private data through pattern selection would be compliant, even when such protection cannot be easily operationalized in technical terms [37]. A stricter interpretation could only certify patterns in which the raw data does not leave the premises, or where more extensive protection exists. Thus, using only **share data** or **delegate** patterns without contracts or additional mechanisms would violate the “*proactive, not reactive, preventative not remedial*” foundation of the principle (p. 5, [38]). We assess our patterns in line with the stricter approach.

6.2. Connections with Related Work

Work connecting privacy, architecture design, and data processing provides two premises to support the use of interaction patterns. First premise is a need for operationalizing privacy-oriented legalese into technically legible steps, and documenting interactions. [39-40] uses privacy engineering methods, empirical validation and PDR to translate GDPR articles into a process model named Protection of Personal Data (ProPerData), sketching out an exemplary pattern candidate for documenting interactions. [29] conclude that despite attempts to satisfy GDPR requirements with privacy-enhancing techniques (e.g., homomorphic encryption or secure

multiparty computation), the gap in processor-provider interactions can only be solved by having rules for documenting them.

Second premise is a need for verifying claims that the proposed architecture is truly privacy-preserving. [5] modelled the threats to which default FL processing is vulnerable, concluding that FL alone cannot satisfy two (data minimization and anonymization) out of three (transparency and consent) core aspects of privacy. To verifiably guarantee protection, FL would need to be combined with additional techniques (e.g., secure enclaves, cryptography, differential privacy), based on trade-offs between accuracy, computational costs and competing objectives.

7. Conclusion

We propose patterns for designing information systems architectures around parties interacting in privacy-preserving contexts. Following a design science methodology, we describe four interaction patterns that involve data processors, providers and third parties, and provide claims as the basis for keeping parties accountable. With an exemplary application, we sketch how patterns can be applied, and discuss how they map to existing privacy principles. The intended benefit of our contribution is to structure the operationalization of privacy principles via a taxonomy of patterns, making preconditions and differences between core and non-core components explicit, and clarifying trade-offs between technical extensions and rule-building activities.

We identify the following limitations. First, preconditions that motivate parties to interact – vulnerability, complementarity, high expected value from collaboration, and high expected costs from non-compliance – might have degrees of interpretation. Our case involved parties with an incentive to improve efficiency, and freedom to establish trust via agreements; other cases may require stricter proofs of satisfied preconditions. Second, secure communication (the fifth precondition) cannot always be expected. Where data, computation specifications or cryptographic keys are vulnerable, patterns may require additional security components. Third, implicit assumptions may be present. We do not specify what policies or contracts should contain, and assume that organizational resistance is minimal. In reality, despite high expected value and low expected risk, data providers may still be reluctant to share data, opting instead for more intensive *isolate* methods or performing no learning at all.

Future work involves three areas. First, another iteration of PDR in different contexts will help operationalize decision-making and explore extensions of privacy-preserving design patterns. We expect that validation will introduce objectives, preconditions, and theories to expand our taxonomy. Second, further specification of tools for assessment and decision-making is needed; we are striving to build a corresponding web tool. Finally, different stages of ML and different assumptions may require different patterns. Combinations with the *specify computation* pattern in the training stage may reveal different ways of instantiating the federated learning concept, depending on other conditions beyond overall trust. We expect the patterns to be a useful conceptualization of privacy-preserving learning outside of our cross-silo scope, and invite research contributions to test the assumption.

Acknowledgements

This research was partially funded by the Bavarian Ministry of Ministry of Economic Affairs, Regional Development and Energy. We thank IBM Research Almaden, IBM Research Ireland and the public city administrations of Munich, Augsburg, and Nuremberg for their cooperation and advice. We thank our reviewers for their careful reading and constructive remarks.

References

- [1] R. Miotto, F. Wang, S. Wang, X. Jiang, J. T. Dudley, Deep learning for healthcare: Review, opportunities and challenges, *Briefings in Bioinformatics* 19.6 (2017) 1236-1246.
- [2] M. Flah, I. Nunez, B. W. Chaabene, M. L. Nehdi. Machine Learning Algorithms in Civil Structural Health Monitoring: A Systematic Review, *Archives of Computational Methods in Engineering* 28.4 (2021) 2621-2643.
- [3] P. Courtiol, C. Maussion, M. Moarii, et al., Deep learning-based classification of mesothelioma improves prediction of patient outcome, *Nature Medicine* 25.10 (2019) 1519-1525.
- [4] B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A., Arcas, Communication-Efficient Learning of Deep Networks from Decentralized Data, *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, AISTATS, 2017*, pp. 1273-1282.
- [5] K. Bonawitz, P. Kairouz, B. McMahan, D. Ramage, Federated Learning and Privacy: Building privacy-preserving systems for machine learning and data science on decentralized data, *Queue* 19.5 (2021) 87-114.
- [6] T. Li, A. K. Sahu, A. Talwalkar, V. Smith, Federated Learning: Challenges, Methods, and Future Directions, *IEEE Signal Processing Magazine* 37.3 (2020) 50-60.
- [7] Q. Li, Z. Wen, Z. Wu, S. Hu, N. Wang, Y. Li, X. Liu, B. He, A Survey on Federated Learning Systems: Vision, Hype and Reality for Data Privacy and Protection, *IEEE Transactions on Knowledge and Data Engineering*, published online, 2021. URL: <https://ieeexplore.ieee.org/document/9599369>
- [8] N. Baracaldo, A. Anwar, M. Purcell, A. Rawat, M. Sinn, B. Altakrouri, D. Balta, M. Sellami, P. Kuhn, U. Schopp, M. Buchinger, Towards an Accountable and Reproducible Federated Learning: A FactSheets Approach, 2022. arXiv: 2202.12443. URL: <https://arxiv.org/abs/2202.12443>
- [9] R. Shokri, M. Stronati, C. Song, V. Shmatikov, Membership Inference Attacks Against Machine Learning Models, *IEEE Symposium on Security and Privacy, 2017*, pp. 3-18.
- [10] M. Fredrikson, S. Jha, T. Ristenpart, Model inversion attacks that exploit confidence information and basic countermeasures, *Proceedings of the 22nd ACM SIGSAC conference on computer and communications security, 2015*, pp. 1322- 1333.
- [11] K. Wei, J. Li, M. Ding, C. Ma, H. H. Yang, F. Farokhi, S. Jin, T. Q. Quek, H. V. Poor, Federated learning with differential privacy: Algorithms and performance analysis, *IEEE Transactions on Information Forensics and Security* 15 (2020) 3454-3469.
- [12] S. Truex, N. Baracaldo, A. Anwar, T. Steinke, H. Ludwig, R. Zhang, Y. Zhou, A Hybrid Approach to Privacy-Preserving Federated Learning, *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security, AISec'19, 2019*, pp. 1-11.
- [13] T. S. Brisimi, R. Chen, T. Mela, A. Olshevsky, I. C. Paschalidis, W. Shi, Federated learning of predictive models from federated electronic health records, *International Journal of Medical Informatics* 112 (2018) 59-67.
- [14] N. G. Packin, RegTech, compliance and technology judgment rule, *Chi.-Kent L. Rev.* 93.193 (2018).
- [15] S. Kacianka, A. Pretschner, Designing accountable systems, *FaccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 424-437, 2021.
- [16] S. Eriksén, Designing for accountability, *NordiCHI '02: Proceedings of the second Nordic conference on Human-computer interaction*, pp. 177-186.
- [17] L. Lessig, *Code and Other Laws of Cyberspace*, Basic Books Inc., New York, NY, 1999.
- [18] L. Lessig, Code is law, *Harvard Magazine*, 2000. URL: <https://www.harvardmagazine.com/2000/01/code-is-law-html>
- [19] M. Buchinger, P. Kuhn, D. Balta, Dimensions of Accountability in Interorganizational Business Processes, *Proceedings of the 55th Hawaii International Conference on System Sciences, HICSS, 2022, Association of IEEE*, pp. 429 - 438.

- [20] M. Brundage, S. Avin, J. Wang, et al., Toward trustworthy AI development: mechanisms for supporting verifiable claims, 2020. arXiv:2004.07213. URL: <https://arxiv.org/abs/2004.07213>
- [21] European Union, Regulation (EU) 2016/679 (General Data Protection Regulation), 2018. URL: <https://gdpr-info.eu/>
- [22] European Union, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final, 2021.
- [23] M. Eßer, P. Kramer, K. von Lewinski, DSGVO BDSG, DatenschutzGrundverordnung, Bundesdatenschutzgesetz und Nebengesetze, 6. Aufl., Köln, 2018.
- [24] A. Toy, D. C. Hay, Privacy auditing standards, *Auditing: A Journal of Practice & Theory* 34.3 (2015) 181-199.
- [25] S. Buckl, F. Matthes, A. W. Schneider, C. M. Schweda, Pattern-based design research – an iterative research method balancing rigor and relevance, in: J. vom Brocke, R. Hekkala, S. Ram, M. Rossi (Eds.), *DESRIST 2013: Design Science at the Intersection of Physical and Virtual Design*, volume 7939 of *Lecture Notes in Computer Science*, Springer-Verlag, Berlin, 2013.
- [26] K. Sedig, P. Parsons, Interaction Design for Complex Cognitive Activities with Visual Representations: A Pattern-Based Approach, *AIS Transactions on Human-Computer Interaction*, 5.2 (2013) 84-133.
- [27] J. O. Horchers, A pattern approach to interaction design, *AI and Society* 15 4 (2001) 359-376.
- [28] N. Truong, K. Sun, S. Wang, F. Guitton, Y. Guo, Privacy preservation in federated learning: An insightful survey from the GDPR perspective, *Computers and Security* 110 (2021) published online.
- [29] T. K., Rodrigues, K. Suto, H. Nishiyama, J. Liu, N. Kato, Machine Learning Meets Computation and Communication Control in Evolving Edge and Cloud: Challenges and Future Perspective, *IEEE Communications Surveys and Tutorials*, 22.1 (2020) 38-67.
- [30] C. Dwork, A. Roth, The algorithmic foundations of differential privacy, *Foundations and Trends in Theoretical Computer Science* 9.3-4 (2014) 211-407.
- [31] X. Yi, R. Paulet, E. Bertino, Homomorphic Encryption, *Homomorphic Encryption and Applications* (2014) 27-46
- [32] C. Marcolla, V. Sucasas, M. Manzano, R. Bassoli, F. H. P. Fitzek, N. Aaraj, Survey on Fully Homomorphic Encryption, Theory, and Applications, *Proceedings of the IEEE* 110.10 (2022) pp. 1572-1609.
- [33] H. Ludwig, N. Baracaldo, G. Thomas, et al., IBM Federated Learning: an Enterprise Framework White Paper, 2020. URL: <https://arxiv.org/abs/2007.10987> (visited on November 15, 2022)
- [34] W. B. Tesfay, J. Serna, K. Rannenber, Privacybot: detecting privacy sensitive information in unstructured texts, *Sixth International Conference on Social Networks Analysis, Management and Security, SNAMS*, 2019.
- [35] X. Zhang, J. Zhao, Y., LeCun, Character-level Convolutional Networks for Text Classification, *Proceedings of the 28th International Conference on Neural Information Processing Systems, NIPS'15*, 2015.
- [36] B.-J., Koops, R. Leenes, Privacy regulation cannot be hardcoded. A critical comment on the 'privacy by design' provision in data-protection law, *International Review of Law, Computers and Technology* 28.2 (2014) 159-171.
- [37] A. Romanou, The necessity of the implementation of Privacy by Design in sectors where data protection concerns arise, *Computer Law and Security Review* 34.1 (2018) 99-110.
- [38] D. Huth, Patterns for GDPR Compliance, *PoEM 2017: Doctoral Consortium and Industry Track Papers*, 2017, pp. 34-40.
- [39] D. Huth, Development of a reference process model for GDPR compliance management based on enterprise architecture, PhD thesis, Technical University of Munich, Munich, Germany, 2021. URL: <https://mediatum.ub.tum.de/doc/1593644/1593644.pdf>

- [40] M. Arnold, R. K. E. Bellamy, M. Hind, et al., FactSheets: Increasing trust in AI services through supplier's declarations of conformity, *IBM Journal of Research and Development* 63.4/5 (2019) 1-13.
- [41] L. Bourtole, V. Chandrasekaran, C. A. Choquette-Choo, H. Jia, A. Travers, B. Zhang, D. Lie, N. Papernot, Machine unlearning, *Proceedings of the IEEE Symposium on Security and Privacy*, 2021, pp. 141–159.

A User-centric View on Data Breach Response Expectations

Felix Hillmann¹, Tim Klauenberg², Lennart Schroeder² and Till Ole Diesterhöft³

¹ University of Paderborn, Warburger Str. 100, Paderborn, 33098, Germany

² University of Göttingen, Platz der Göttinger Sieben 5, Göttingen, 37073, Germany

³ University of Göttingen, Humboldtallee 3, Göttingen, 37073, Germany

Abstract

Due to the growing prevalence of data breaches and the associated negative outcomes, data breaches pose a serious problem for companies. Since universal response strategies may not fully address diverse customer expectations, their effectiveness could be limited. As a result, understanding customer expectations serves as the cornerstone of a successful response strategy. By integrating prior data breach research with expectation confirmation theory, we examine individual customer expectations across a wide range of situations and business environments. Therefore, we conducted twelve qualitative interviews. Our findings enrich the body of research on data breaches by highlighting the individualized nature of customer expectations regarding data breach responses, which are shaped by numerous factors. We also discuss our contributions to the literature and the implications for managing data breach responses more effectively.

Keywords

Data breach response, customer expectations, expectation confirmation theory

1. Introduction

According to the Ponemon Institute [21] for 83% of the companies it is not a question of if, but when a data breach will happen. Companies that store large amounts of personal data face a high risk of data breaches [10, 14], which can have various negative effects. Companies affected by breaches must inform affected customers and regulatory authorities [23]. Therefore, the importance of a cost-effective communication and response strategy that meets customer expectations is increasing [16]. Such a strategy aims to minimize damage to the company [36] and mitigate the negative impact caused by disgruntled customers [18, 27]. Various response strategies have been analyzed in the literature as recovery actions. Compensation and apology have been identified as common practices in addressing data breaches [16, 18]. Although Goode et al. [16] and Hoehle et al. [18] have shown that the success of the company's response strategy strongly depends on customer expectations. Consequently, companies need to ascertain customer expectations and incorporate them into their respective response strategy to minimize the negative impacts of a data breach [16, 36]. However, current literature has yet to explore the diversity of customer expectations in a proactive and qualitative approach. The Expectation Confirmation Theory (ECT), proposed by Oliver [38], supports an understanding of the importance of aligning the response strategy with individual customer expectations, impacting overall satisfaction and trust in the company. Given this background, our study aims to answer the following research question (RQ):


RQ: What are customers' expectations of a company's response to a data breach?

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ felixhi@campus.uni-paderborn.de (F. Hillmann); tim.klauenberg@stud.uni-goettingen.de (T. Klauenberg); lennart.schroeder02@stud.uni-goettingen.de (L. Schroeder); tillole.diesterhoeft@uni-goettingen.de (T. O. Diesterhöft)

ORCID: 0009-0006-0129-2000 (F. Hillmann); 0009-0007-0376-5374 (T. Klauenberg); 0009-0003-2315-2525 (L. Schroeder); 0000-0002-4141-3261 (T. O. Diesterhöft)

© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

Drawing on previous research in data breach response expectations and ECT, we examine the alignment between companies' response strategies and individual customer expectations [16, 18, 36].

To answer the RQ, we conducted twelve qualitative interviews with affected or potentially impacted customers. These interviews explored various customer expectations that have not been previously studied. The identified expectations can be further examined for their effectiveness in response to data breaches. Additionally, these findings have practical implications as companies can optimize their data breach response strategy based on individual customer expectations. Thus, this study specifically targets company security management.

2. Research Background

2.1. Data Breaches

A data breach refers to the unauthorized use, storage, processing, or disclosure of personal data in violation of data protection laws, which can cause harm to individuals, companies, or governments [37, 39]. Breaches can occur through various means, such as data loss or theft, hacking, unauthorized access, accidental disclosure, lack of security measures, or abuse of personal data [4, 24, 32, 52].

Data breaches have become a common and serious threat due to increased reliance on digital technology and the internet [35, 40]. Despite increased cybersecurity awareness and investment, companies continue to struggle with securing their networks and data, resulting in rising costs of data breaches [22]. No company is immune to attacks or breaches, whether intentional or due to human error [16, 49].

Data breaches pose significant threats to privacy and security, particularly when sensitive personal information is involved [31]. Laws and regulations have been enacted in many countries to protect personal data and hold companies accountable for breaches [23]. The impact of data breaches on affected customers can include identity theft and financial losses [43]. Customers may lose trust in the company that experienced the breach, leading to a decline in customer loyalty and a loss of business for the company [5, 34, 36]. Companies may face financial losses, legal penalties, reputation damage, and decline in sales [25, 47]. Recovering from a breach requires significant investments in IT infrastructure, employee training, and preventive measures [7, 13, 22]. Overall, the consequences of a data breach can be far-reaching and affect not only the company but also its customers.

2.2. Review of Data Breach Response Strategies Research

To prevent data subjects from being harmed due to improper data disclosure [31], laws are being enacted that require companies to notify affected customers in the event of a data breach [23]. In this context, it has been shown that the challenge is to adapt the company's response strategy to the affected customer expectations [16]. The majority of companies strategically employ apologies and compensation, which previous research on data breach response has found to have a positive impact on perceived service quality, customer loyalty, and repurchase intent, thus minimizing the damage done [16, 18]. Regardless, companies often experience a significant rate of customer attrition due to the discrepancy between their response strategy and customers' expectations [16, 18]. Additionally, many companies opt to initiate external disclosure of a data breach only after they have gained a sufficient understanding of the breach and conducted a thorough investigation [28]. Regrettably, delays caused by a lack of response plans can result in ineffective and prolonged communication with customers, leading to decreased customer satisfaction [48]. Although companies may provide financial compensation, such as free products or services, discounts, or credit monitoring, to customers affected by a data breach, as well as communicate with them about the incident, offer an apology, and provide details on the breach and how to protect oneself [17], there is uncertainty about how to properly align compensation

levels with customer expectations [15]. Determining the appropriate level of compensation is a challenging and costly process [18]. Any deviation from customer expectations, whether exceeding or falling short, may result in reduced satisfaction and repurchase intentions [16]. Moreover, the severity of a data breach can vary [35, 42], affecting customer reactions and expectations differently, which necessitates a careful balance between compensation and severity to meet customer expectations without overcompensating. In conclusion, managers must strive to match compensation with customer expectations to ensure future customer retention in the event of a data breach [16, 36]. Consequently, there is a growing need to expand research aimed at meeting customer expectations. Focusing on these issues can help companies mitigate the negative impact of data breaches and strengthen their relationships with customers.

2.3. Expectation Confirmation Theory

The expectation confirmation theory (ECT) is a widely studied theoretical model in the field of consumer behavior and was originally proposed by Oliver [38] to explore the concept of customer satisfaction. Based on this theory, individuals pre-establish their expectations regarding a product or service before engaging with it, and subsequently assess their level of satisfaction based on the degree to which the product or service meets or surpasses those initial expectations [1, 38]. If a product or service satisfies or surpasses predetermined expectations, the individual experiences confirmation, resulting in positive satisfaction. Conversely, if the product or service fails to meet predetermined expectations, the individual experiences disappointment, leading to negative satisfaction [3]. Furthermore, the theory posits that post-consumption behavior is influenced by cognitive dissonance, a psychological state of mental discomfort that arises when individuals hold conflicting beliefs or values. A significant discrepancy between an individual's expectations and their actual experience with a product or service is likely to result in cognitive dissonance [38]. Research has demonstrated that ECT can be applied to multiple domains, including product repurchase [44], healthcare [8] and e-commerce [33]. Given the demonstrated predictive power of ECT in the various domains, we believe it is appropriate to use ECT to examine customer behavior on a company's data breach response strategy. In the course of a data breach, preserving customer loyalty is a crucial factor in a company's long-term costs [36], making it essential for companies to meet consumer expectations regarding their response to the breach. Nonetheless, there remains a paucity of research regarding the customer's viewpoint of response strategies and their expectations in this regard. According to ECT, companies should conduct a thorough exploration of customer expectations to align their response strategy and meet customer expectations following a data breach. This proactive approach can provide a useful way for companies to gain detailed insights into customer expectations, enabling them to adjust their response strategies and mitigate potential negative impacts on customer satisfaction, retention, and churn [16, 36]. Furthermore, this approach can assist companies in adapting their response strategies according to the diverse levels of severity inherent in various data breaches, as well as in gaining a comprehensive understanding of customers' distinctive expectations associated with each type of breach. To identify and gain an overview of these diverse and individual expectations regarding response strategies, we are conducting a qualitative study.

3. Research Methodology

Qualitative research places a significant emphasis on the lifeworld of individuals, aiming to comprehend specific perspectives [12]. This approach focuses on the subjective experiences of those involved [2]. Qualitative research describes social phenomena in detail and depth, allowing for a more nuanced understanding of human experience and behavior [26]. Since this research focuses on a user-centered view of data breach response expectations, qualitative research is appropriate for conducting this project. Therefore, the framework of Kuckartz & Rädiker [29] will be used in this thesis as it focuses on conducting a qualitative content analysis based on interviews. Fundamentally, it is about subjectivity, as Flick [11] points out, and the related

elicitation of the experiences and perceptions of those affected [2]. This can be achieved through qualitative social research.

3.1. Data Collection

In order to capture the user-centered view of expectations in response to data breaches, the problem-centered interview according to Witzel [50] was chosen. The problem-centered interview is a semi-structured open questioning method that focuses on a problem yet still allows the interviewees to express their personal viewpoints relatively freely [19]. The problem-centered interview is well-suited to the project of this study, which focuses on a user-centered view of expectations in response to data breaches. This topic represents a significant social problem of the modern age that affects both customers and companies. As previously noted, current measures such as compensation and apology often fail to meet customers' expectations of an appropriate response strategy and are insufficient in terms of recovery compared to service failures. Consequently, these measures cannot be fully applied in response to data breaches, and they do not necessarily provide full compensation for any damage incurred [9, 18]. Thus, it is crucial to identify a suitable response strategy that better meets customer expectations and strengthens the relationship between company and customer. Given the complexity of this topic, guiding the interviewer through targeted and follow-up questions during the problem-centered interview can yield the most nuanced and comprehensive data possible. For these reasons, the problem-centered interview was chosen. First, a short questionnaire was created using Qualtrics software to capture the socioeconomic background of the respondents, as suggested by Witzel [50]. This information also serves to enable the interviewer to prepare appropriately for the interview. Participants are asked about the frequency and companies involved in any past data breaches they have experienced, in order to address these cases specifically during the interview. If participants have not experienced a data breach, they were asked to provide the social media platforms they use and the health insurance company they are insured with, so that a representative fictional scenario can be presented to them. Furthermore, an interview guide was developed to serve as a frame of reference, incorporating pre-written questions that cover various topics [50]. The questions are designed to assess customer expectations following a data breach and are thus tailored to answer the research question. If the participants have not experienced a data breach, the guide includes a personalized scenario based on the information provided in the questionnaire. This approach is intended to ensure that all participants can best empathize with the case that they are affected by a data breach. On the other hand, if the participants have already been affected by a data breach, actual cases are addressed during the interview. This reference to real cases should allow to obtain valuable information about the expectations and the actual reaction of the companies. A total of twelve participants were recruited for the interviews. The demographics of the participants (age, gender, education level) were considered to ensure a diverse sample. The questionnaire revealed that four participants had experienced a data breach, five were unaware, and three had not yet encountered such an incident. The interviews lasted an average of 25 minutes, ranging from 14 to 39 minutes and were conducted between January and February 2023. The data collection process was concluded when it was determined that no new significant information was being revealed, thus achieving theoretical saturation, and ensuring that no additional properties or dimensions would emerge during the analysis [45]. All interviews were recorded and transcribed, following the transcription guidelines set forth by Kuckartz & Rädiker [29]. The transcription process resulted in a total of 99 pages of material.

3.2. Data Analysis

Since the content-structuring qualitative content analysis according to Kuckartz & Rädiker [29] is used in this work, the following explanations should make transparent how the results of this study were obtained.

Phase 1 - Initiating text work, memos, case summaries: The text was reviewed for components essential to answering the research question, and comments and notes were added.

Phase 2 - Develop main categories: The focus in this phase is on developing the main categories. During this phase, "Customer Expectations of the Company's Respond" could be identified as the first central main category.

Phase 3 - Coding data of the main categories (1st coding process): Any text passages with expectations were assigned the main category "Customer Expectations of the Company's Response" accordingly. If new main categories could be identified, they were included in addition to this one. It should be noted that text passages or individual sentences need not be assigned to a single category exclusively. A passage can pertain to several categories if multiple topics are addressed. The data were coded by two researchers. To ensure consistency, the coding results were reviewed and discussed together after every three interviews analyzed.

Phase 4 - Forming inductive subcategories: The next phase in the content analysis process is the differentiation of main categories into more specific subcategories. In this step, the expectations and thus the main category "Expectation of the company" were transferred into the concrete expectations.

Phase 5 - Coding data with subcategories (2nd coding process): In a second coding process, all text passages previously identified only as an expectation were coded with the appropriate subcategory and thus with the specific expectation. Analogous to the procedure in phase 3, further subcategories were included if they were identified.

Phase 6 - Simple and complex analyses: The sixth phase of this process involves preparing the presentation of the research results. Thus, all categories were examined, and interrelationships were explored in order to answer the research question and, beyond that, to possibly arrive at further findings.

Phase 7 - Writing down results and documenting procedures: This step reflects the elaboration of the present study.

4. Findings

4.1. Customer Expectations of the Company's Response

In terms of the RQ, customers' expectations of company response form the main category of this research. Table 1 illustrates how respondents' statements were assigned to each subcategory. This includes the various expectations that respondents have expressed regarding the company response because of the data breach (see Appendix, Table 4). To ensure anonymization of respondents, ID's B1 through B12 are used in the following.

Table 1
Categories of customers' expectations of the company's response

Category	Category definition	Example	Coding Rule
Compensation N = 10	The category compensation includes all customer expectations of compensation from the company as a result of a data breach.	"With the negligence one, there definitely, I do expect compensation at the end."	Applies if respondents expect compensation, even if it is not explicitly stated in what form this compensation is expected.
Notification N = 12	Includes all statements in which customers want to be informed about the incident.	"[I] would also want to know how it came about now."	Applies in the case that the respondents should be informed about the data breach.

Table 2
Categories of customers' expectations of the company's response (continued)

Category	Category definition	Example	Coding Rule
Follow-up Notification N = 5	Contains all statements in which further funding was expected from the company beyond initial information.	"I would always like to be informed about the next steps and that we might be able to talk a bit about this."	Applies when respondents want to be kept informed and want to know what happened with the data breach. Also, if this is mentioned in a different context.
Fast reaction of the company N = 6	Includes all statements that expect the company to respond quickly as a result of a data breach.	"Act as fast as possible, certainly, and contact affected parties and get it fixed as soon as possible."	Applies when respondents' statements call for or expect the company to respond quickly.
Transparency N = 5	Includes all statements in which companies are expected to be transparent in their dealings with data subjects.	"The more sensitive the data becomes, the more important it is (...) that companies are transparent (...)"	Applies if in the statements of the respondents a transparent handling of the data breach of the companies with the customers is expected.
Apology N = 9	The apology category includes all statements in which an apology was expected or requested.	"A company can also apologize here only I think in writing personally to one."	Applies when respondents expect an apology from the company.
Empathy N = 5	Includes all statements in which respondents expected empathy in communicating, apologizing, or communicating with the company.	"As long as (...) an empathic apology comes for it."	Applies when respondents expect empathy from the company.
Measures N = 12	Includes all statements in which respondents expected the company to take measures as a result of the data breach.	"That they'll make sure it never happens again."	Applies if the statements contain concrete suggestions for improving safety or refer to the fact that the problem will be remedied, and this will no longer occur.
Support N = 8	Includes all statements in which assistance is expected to be provided to affected individuals in dealing with a data breach.	"What consequences, what I could have to fear, how you would advise me, how I should best proceed regarding my data breaches."	Applies when the company is expected to take a collaborative approach to assist in dealing with the incident and to provide information about possible consequences and risks following a data breach.

Table 3
Categories of customers' expectations of the company's response (continued)

Category	Category definition	Example	Coding Rule
Participation in the decision-making process N = 6	Includes all statements where affected individuals want to be involved in the company's response process.	"I would like to find a way to satisfy both parties (...) [and] would like to participate in the decision-making process."	Applies when respondents want to be involved in the solution and decision-making process and can actively contribute their opinions.

Compensation: Ten out of twelve respondents expressed the expectation of compensation. The interviewees have different expectations and demands regarding the format and amount of compensation. In addition, it was also mentioned that there are different factors that influence the expectation of compensation as a response. In addition, two central subcategories of compensation were identified: Financial compensation and free/discounted services. Interviewee B3 stated that companies are only expected to pay compensation if the data breach has caused damage to the customer. If no harm has occurred, compensation is not necessarily expected, but is still perceived as positive. Furthermore, B5 has additionally mentioned that compensation is explicitly expected if the company has acted negligently. Connecting to this, B10 said that high compensation is expected in particular if sensitive data has been published. If the Severity of the data breach is less, compensation is also expected to be less. In addition to the general expectation of compensation, the expectation of financial compensation was also identified during the interviews. This category is defined by the explicit expectation of financial compensation expressed by the interviewees. In total, the expectation of financial compensation was expressed by eight interviewees. Within the financial compensation, this reveals that the expectation of financial compensation is influenced by the severity of the data breach. To this, it was also expressed by B4 that financial compensation is expected when damage has occurred to the respondents. In addition to the severity of the damage, B3 said that the type of data is a factor influencing the expectation of financial compensation, especially when sensitive data is involved. One further subcategory of compensation is the expectation of free/discounted services. This subcategory includes paragraphs in which the expectation of free or discounted services was mentioned. Free/discounted services were expressed by two interviewees. It was mentioned by B1 that the service should be suitable for the company and a service offered should be free or at a reduced price.

Notification: Interviewee B11 primarily expect to be notified about the breach and receive an explanation of how it happened and its potential causes. In addition, Respondent B5 and B8 suggested providing regular updates on the investigation's status, which should include information on the cause, scope, and impact of the breach, as well as which data was stolen and the extent of individual impact.

Follow-up Notification: Five out of twelve participants expressed a desire for follow-up notifications in addition to the initial notification. They seek information on the details of the data breach, the measures being taken, and preventive measures for future incidents. For example, B10 would like to be informed about the outcome of the data breach.

Fast Reaction of the Company: Half of the interviewees mentioned that they expect prompt action from the company in response to a data breach. This expectation is addressed on the one hand to the notification of this incident, and on the other hand to the measures that should be made in consequence, as B1 noted. Respondent B4 also expressed this expectation in the event of a data breach being reported through the media before being acknowledged by the company. In addition, B6 expressed that timely notification is expected especially when sensitive data is involved (see Type of data).

Transparency: Transparency describes the extent to which information is visible and accessible [51]. Five respondents expect companies to be transparent in their dealings with customers. In the case of sensitive data, respondent B3 expects increased transparency regarding

the whereabouts of the data and the company's-initiated measures. Furthermore, B5 expected continuous updates from companies during longer investigations, which should include the status of the investigations and future measures or precautions to be taken, as well as final results or findings. In addition, B5 mentioned the reduction of uncertainties and fears as a possible consequence of increased transparency.

Apology: Within the interviews, nine out of twelve interviewees expressed the expectation of an apology. In this context, different conditions were expressed when an apology is expected. For instance, Respondent B4 stated that an apology is only expected if the company is responsible for causing the data breach (see Company fault). Additionally, negative reactions may occur if the company does not apologize and does not meet expectations, as B7 stated in this context. B5, in turn, expects an apology regardless of whether the company is to blame for the data breach. However, in addition to the terms of when compensation is expected, there are also different expectations in which manner the apology should be delivered. In this context B1, expressed that a written apology was expected.

Empathy: The category of empathy is characterized by respondents' expectation of empathy in communication with the company as a result of a data breach. During the interviews, five interviewees said the expectation of empathy when communicating, apologizing, or communicating with the company. Thus, B7 expressed that empathy in the communication increases customer's forbearance as long as an empathetic apology is provided. Furthermore, it was added that the increased use of empathy is perceived positively and increases the customer's comprehension. Complementing this, B7 additionally specified that an empathic apology is expected.

Measures: All interviewees expected the company to perform measures in consequence of the data breach. In this context, this refers to all statements that contain specific suggestions for improving security or refer to the fact that the problem will be remedied. As several participants, including B4, noted, the measures should ensure that data breaches do not recur (see). A wide variety of possibilities are mentioned for companies to avoid such incidents as well as to minimize the damage afterwards. Specific ways to realize this expectation and avoid incidents of this nature were explained by B2 and B4. B8 additionally states that depending on the type of data, a higher level of data protection is expected.

Support: B10 expects support in dealing with a data breach and that the company will take a collaborative approach and provide information about possible consequences and risks following a data breach. Furthermore, the respondents also mentioned that they would like to receive preventive and protective measures or a possible guide on what to do or recommendations for action as noted by B4. This is supported by B8 and B9.

Participation in the Decision-making Process: When involving affected individuals in the company's response process to data breaches, half of the respondents expect to be included in the decision-making and solution-finding processes of the company's response. This is exemplified by B1's statement. Furthermore, B4 suggested that companies should offer different compensation options. Nevertheless, the opinion and input of affected individuals should be given room to maneuver, as B2 pointed out.

4.2. Influencing Factors of Customer Expectation

In addition to customer expectations, the interviews also identified various factors that influence customer expectations. Table 2 illustrates how respondents' statements were assigned to each subcategory. Although these have already been mentioned in the category of expectations, they are presented in their entirety in the following category (see Appendix, Table 5).

Table 2
Categories of influencing factors of customer expectation

Category	Category definition	Example	Coding Rule
Severity of the data breach N = 12	Includes all statements in which the severity of the data breach had an impact on the expectation.	"(...) and depending on the severity, of course, you would have to see whether financial compensation or compensation in kind would come into question."	Applies if the expectations regarding the severity of the data breach change. The same applies if they do not change and this factor is explicitly mentioned in this context.
Type of company N = 9	The type of company as a subcategory describes the influence of the type of company on the customer's expectation of the company's response as a consequence of a data breach.	"Because of the size of the company, I would definitely expect them to be more transparent about it, to follow up when a data breach happens."	Applies when the type of company had an impact on expectations
Type of Data N = 12	This category is characterized by respondents having different expectations for different types of data.	"Yes, if more sensitive data is affected, I expect fast notification."	Applies when the type of data has an impact on expectations.
Personal responsibility of the customer N = 4	This influencing factor affects expectations when customers themselves are responsible for what data and information they share.	"If it's kind of your own fault that something like that happens, that you change passwords more often or email addresses or something."	Applies when personal responsibility has had an impact on expectations.
Company fault N = 4	This category is defined by the fact that fault by the company affects expectations.	"Yes, so if I clicked on some phishing email and then my data was stolen, then of course I don't expect compensation from the company. (...) So if it's clear that the company has nothing to do with it, then I don't expect an apology."	Applies when expectations change, when the company is at fault or the not at fault.

Severity of the Data Breach: The respondents stated that the severity of the abuse, especially due to the type of data (see Type of data) and potential consequences, has an impact on their expectations of the company and customer reactions. B1 and B10 consider categorizing sensitive data as more severe and leaving an company in the event of far-reaching consequences. Furthermore, B3 mentioned that with increasing severity of the data breach, higher transparency, and more information from the company regarding the violation are expected.

Type of Company: The type of company describes the influence of the type of company on the customer's expectation of the company's response as a consequence of a data breach. In this context, nine of the twelve respondents expressed that the type of company had an influence on their expectations. In this regard, expectations are higher for companies that collect sensitive data than for companies with less sensitive data, as B7 said in this regard. In addition, interviewees B4 described that they have higher expectations regarding transparency and notification for data breach response at larger companies. It is also described that at smaller companies there is a higher level of understanding when a data breach occurs, as also noted by B4.

Type of Data: All respondents expressed that the type of data has an influence on their expectations towards the company. Firstly, it was expressed by B11 that the breach of sensitive data, especially health data is perceived as more significant. Furthermore, interviewees stated an expectation of compensation, in particular for sensitive data (see Compensation). In addition to compensation, more transparency in the handling of data breach was expected, especially for sensitive data (see Transparency). In addition, B6 expected rapid notification and response from the company. In addition, it was expressed by B11 that there is a higher expectation of protective measures by the company for sensitive data. On the other hand, B2 and B3 mentioned that for less sensitive data, compensation in the same way is sufficient regardless of the severity of the data breach (see Severity of data breach), as long as no personal or irreversible damage occurs.

Personal Responsibility of the Customer: In the context of this research, personal responsibility of the customers could be identified as an influencing factor. For four participants, this category influences expectations as a result of a data breach. It affects expectations if customers themselves are responsible for what data and information they share, as B7 notes. In addition to lower expectations to the company, the customer reaction is also influenced. This is changed by customers taking personal responsibility to protect themselves by taking measures, B4 commented.

Company Fault: The company fault is another factor influencing expectations. It was said by B4 that no compensation or apology is expected if the company is not at fault. If the company is at fault, then the company is expected to approach the customer with an optimal solution and participation in the decision-making process is therefore rejected, as B5 noted. As B5 also said, expectations are higher when the company is at fault, and in that case expects notification, information about what measures will be taken, an apology, and compensation. In addition, financial compensation in this context is seen as positive by B2.

4.3. Meeting the Customer Expectations

In connection with the expectations, we were able to identify various statements in the interviews that provide information about the reaction to meeting or not meeting the customer expectations. This category therefore shows, the perception of the customers, the impact on their attitude towards the company and what steps they would adopt in these cases. Table 3 illustrates how respondents' statements were assigned to this category. All respondents in this category indicated that not meeting expectations has negative consequences for the customer- company relationship (see Appendix, Table 6).

Table 3
Meeting the customer expectation

Category	Category definition	Example	Coding Rule
Meeting the customer expectation N = 7	The category contains all statements about what happens if expectations are met or not met.	“Because you're disappointed, simply. You had expectations, they are not fulfilled or just destroyed. And yes, then I am sad and that is normal.”	Applies when respondents gave their assessments of meeting and not meeting expectations.

B6 expressed disappointment in this regard. If expectations are not met, the majority of respondents indicate that they would leave the company or potentially switch to another one, such as B2 stated. If expectations are met, however, respondents indicate that their opinion and intention to enter into a business relationship with the company is reinforced, as B2 also stated. In addition, B8 stated that meeting expectations can strengthen the relationship and trust with the company. Respondents did not indicate whether they would also consider failure to meet expectations, in the sense of exceeding expectations, to be negative.

5. Discussion

Based on literature-based knowledge, twelve qualitative interviews were conducted under consideration of the ECT to gain an overview of the strongly individualized expectations [18, 36] and needs of customers in different situational and company contexts. The interviews conducted in this study enabled the interviewees to express their expectations of companies' response strategies in the event of a data breach.

5.1. Contribution to Literature

Our study contributes to the literature on data breaches and service failure in multiple ways, advancing existing research. This study employs a proactive and qualitative methodology to gather and analyze data on customer expectations and influencing factors before the occurrence of a data breach. We present valuable insights into individualized expectations of different demographic groups and contribute to the security literature by presenting unique data and theoretical perspectives, enabling companies to develop adjustments and novel approaches to meet customer expectations following a data breach. In doing so, our findings support and extend the current research of Goode et al. [16] and Hoehle et al. [18], which emphasize that companies should align their actions with customer expectations. Consistent with previous findings from the literature and research on the expectation of apology [6, 36] and compensation [16, 18, 30], respondents in our interviews also expressed this expectation. Additionally, they expressed a desire for companies to involve affected parties in the response process, which supports the earlier research findings of Diesterhöft et al. [9]. In addition to corroborating expectations derived from existing literature, this study yielded novel findings. These include new expectations as well as associated influencing factors. While our findings confirm previous research indicating that customer trust and loyalty are significantly impacted by a data breach, we found that a well-tailored response strategy that meets customer expectations can mitigate the loss of trust and loyalty. Moreover, the results of the interviews indicate that those affected expect greater empathy, transparency, and follow-up notifications from the company. These findings are consistent with crisis management research, which advocates open and continuous communication for companies [20, 46]. Another aspect of our study encompasses the identification of diverse factors influencing customer expectations pertaining to a company's response strategy in the context of data breaches. These factors include the type of data affected, the company's characteristics, the extent of the customer's culpability, the degree of the company's responsibility, and the severity or scope of the data breach incident. Our study partially contradicts existing literature and shows that affected individuals expect communicative interactions, such as follow-up notifications regarding the current status or updates, in order to stay informed. This insight introduces a novel approach wherein active, continuous communication in data breach response should be considered more as an ongoing process. This perspective carries implications for prior experiments and research designs, potentially leading to a reevaluation of response strategy effects on customers, subsequently altering their responses and perceptions.

5.2. Implications for the Management of Data Breach Responses

In addition to the theoretical implications, the results of this work also provide practical implications for companies to potentially improve the management of data breach responses. First, we were able to identify various customer expectations that concern how companies communicate with customers. These can be implemented as part of the legal notification of the data breach to optimize it in terms of customer expectations [23]. Specifically, when sensitive data is breached, respondents expect the company to communicate as soon as possible, as well as transparency in communication and handling of the data breach. In this context, interviewees also expect the company to keep the customer informed of the further progress of the data breach by means of follow-up notifications. In addition, our research provides new insights for the practical implementation of apology, which have not been previously considered in the literature [16, 18]. Our research suggests that apologies are particularly expected when the company is definitely at fault for the occurrence of the data breach. Moreover, in the context of communicating the data breach, it was also expressed that empathy is expected. This insight can be adopted by companies in the context of apology and notification in order to increase customer satisfaction and reduce the negative consequences of a data breach. Second, our research extends the outcomes-based approach of compensation [16]. Our research suggests that there are different expectations regarding the sensitivity of data and the severity of individual consequences related to compensation. For these cases, the majority of respondents indicated higher expectations for the level of compensation. Therefore, companies must manage data breach response in a situation-specific manner, considering the individual customer's expectations. Third, the interviews indicate several actions that respondents expect to take as a result of a data breach to address the problem. In particular, companies that hold sensitive data are expected to take preventive measures to minimize the likelihood of data breaches so that the company is better protected in the future.

5.3. Limitations and Future Research Directions

It is important to acknowledge that the findings of our study are subject to certain limitations stemming from the reliance on information and insights derived from the participants' experiences, opinions, and attitudes. Consequently, generalization of the results may not be feasible. First, it must be considered that only four of the twelve interviewees had previously been affected by a data breach. Thus, the expectations were only expressed in the context of a fictional scenario. As a result, it may not always match the actual response and expectations of a real data breach [16]. Second, the interviewees who were already affected by a data breach did not express their expectations immediately after the incident of the breach, but rather after the breach had occurred. Thus, their expectations could be biased by the time gap. Conducting interviews with affected parties immediately after the incidence of a data breach can increase representativeness. Third, it is crucial to examine the limitations associated with the experiment, particularly concerning the sample size and representativeness of the participants. In cases where the sample size is small, the results may lack generalizability and make it difficult to detect general trends or patterns in the results. Therefore, in future research also the sample size should be increased to achieve higher representativeness. Notwithstanding these limitations, qualitative research employing interviews can prove to be an invaluable method for delving into the experiences, opinions, and attitudes of the participants. It is of utmost importance to acknowledge the limitations, while appropriately interpreting and presenting the results to ensure the credibility and reliability of the study's conclusions. To strengthen our findings, future research could use a quantitative methodology. In the context of influencing factors, prospective research could be conducted. This would allow an analysis of customer expectations regarding the level of compensation and avoid the associated uncertainty as to how these can be reconciled [15].

6. Conclusion

Building upon prior research on data breaches and ECT, this study aims to investigate the expectations customers hold regarding a company's response to a data breach. We conducted problem-centric interviews within a qualitative study (n=12) to obtain an overview of individual customer expectations. Our research implies that customer expectations are highly personalized and influenced by various factors. In this regard, we lay the groundwork for future research to quantitatively examine additional expectations and consider influencing factors in the study design. Consequently, our research offers novel insights that should be taken into consideration when designing future research experiments. Moreover, we contribute valuable knowledge to practitioners by emphasizing the importance for companies to understand and be aware of customer expectations. Companies should tailor their data breach response to the specific situation, taking into account the expectations of individual customers. This highlights that the research area possesses additional gaps that warrant exploration in future studies. By examining the diverse and individual expectations of affected parties concerning the response strategies employed by companies during a data breach, the findings of this study have already made a substantial contribution in addressing these gaps.

References

- [1] Anderson, E.W. & Sullivan, M.W. (1993), 'The Antecedents and Consequences of Customer Satisfaction for Firms', *Marketing Science*, 12(2), pp. 125–143.
- [2] Blumer, H. (1980), 'Der methodologische Standort des symbolischen Interaktionsismus', in Arbeitsgruppe Bielefelder Soziologen (ed.) *Alltagswissen, Interaktion und Gesellschaftliche Wirklichkeit*. [Online]. Wiesbaden: VS Verlag für Sozialwissenschaften. pp. 80–146.
- [3] Bolton, R.N. & Drew, J.H. (1991), 'A Multistage Model of Customers' Assessments of Service Quality and Value', *Journal of Consumer Research*, 17(4), pp. 375–384.
- [4] Boss, S.R., Kirsch, L.J., Angermeier, I., Shingler, R.A. & Boss, R.W. (2009), 'If someone is watching, I'll do what I'm asked: mandatoriness, control, and information security', *European Journal of Information Systems*, 18(2), pp. 151–164.
- [5] Cavusoglu, H., Mishra, B. & Raghunathan, S. (2004), 'The Effect of Internet Security Breach Announcements on Market Value: Capital Market Reactions for Breached Firms and Internet Security Developers', *International Journal of Electronic Commerce*, 9(1), pp. 70–104.
- [6] Chan, E.Y. & Palmeira, M. (2021), 'Political ideology moderates consumer response to brand crisis apologies for data breaches', *Computers in Human Behavior*, 121p. 106801.
- [7] Cheng, L., Liu, F. & Yao, D.D. (2017), 'Enterprise data breach: causes, challenges, prevention, and future directions: Enterprise data breach', *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 7(5), pp. 1–14.
- [8] Chou, H.-K., Lin, I.-C., Woung, L.-C. & Tsai, M.-T. (2012), 'Engagement in E-Learning Opportunities: An Empirical Study on Patient Education using Expectation Confirmation Theory', *Journal of Medical Systems*, 36(3), pp. 1697–1706.
- [9] Diesterhöft, T.O., Schweneker, S.I., Masuch, K., Aslan, A. & Braun, M. (2022), 'The Role of Uncertainty in Data Breach Response Processes - A Reactance Theory Perspective', in 'Forty-Third International Conference on Information Systems', pp. 1–17.
- [10] Edwards, B., Hofmeyr, S. & Forrest, S. (2016), 'Hype and heavy tails: A closer look at data breaches', *Journal of Cybersecurity*, 2(1), pp. 3–14.
- [11] Flick, U. (2010), 'Gütekriterien qualitativer Forschung', in Günter Mey & Katja Mruck (eds.) *Handbuch qualitative Forschung in der Psychologie*. 1. Aufl [Online]. Wiesbaden: VS Verlag für Sozialwissenschaften. pp. 395–407.
- [12] Flick, U., Kardorff, E. von & Steinke, I. (2015), in *Qualitative Forschung. Ein Handbuch*. 11th edition. Rowohlt Taschenbuch.

- [13] Forbes R2_11 (2014), 'NVIDIA Corporate Network Breached', Forbes Media LLC., <https://www.forbes.com/sites/davelewis/2014/12/29/nvidia-corporate-networkbreached/?sh=489544a36241>. Accessed 18.03.2023.
- [14] Gatzlaff, K.M. & McCullough, K.A. (2010), 'The Effect of Data Breaches on Shareholder Wealth', *Risk Management and Insurance Review*, 13(1), pp. 61–83.
- [15] Gelbrich, K. (2010), 'Anger, frustration, and helplessness after service failure: coping strategies and effective informational support', *Journal of the Academy of Marketing Science*, 38(5), pp. 567–585.
- [16] Goode, S., Hoehle, H., Venkatesh, V. & Brown, S.A. (2017), 'User Compensation as a Data Breach Recovery Action: An Investigation of the Sony PlayStation Network Breach', *MIS Quarterly*, 41(3), pp. 703–727.
- [17] HHS (2013), 'Breach Notification Rule', U.S. Department of Health & Human Services, <https://www.hhs.gov/hipaa/for-professionals/breach-notification/index.html>. Accessed: 18.03.2023.
- [18] Hoehle, H., Venkatesh, V., Brown, S., Tepper, B. & Kude, T. (2022), 'Impact of Customer Compensation Strategies on Outcomes and the Mediating Role of Justice Perceptions: A Longitudinal Study of Target's Data Breach', *MIS Quarterly*, 46(1), pp. 299–340.
- [19] Hölzl, E. (1994) 'Qualitatives Interview', in Arbeitskreis Qualitative Sozialforschung & Otmar Chorherr (eds.) *Verführung zum Qualitativen Forschen: Eine Methodenauswahl*. [Online]. Wien: . pp. 61–68.
- [20] Huang, Y.-H. & Su, S.-H. (2009), 'Determinants of consistent, timely, and active responses in corporate crises', *Public Relations Review*, 35(1), pp. 7–17.
- [21] IBM & Ponemon Institute (2022), 'Cost of a data breach 2022 - A million-dollar race to detect and respond', <https://www.ibm.com/reports/data-breach>. Accessed 18.03.2023.
- [22] IBM & Ponemon Institute (2020), 'Cost of a Data Breach Report 2020',
- [23] Identity Theft Resource Center (2020), 'Data Breach Report in 2018', <https://www.idtheftcenter.org/post/identity-theft-resource-centers-annual-end-of->, Accessed: 18.03.2023.
- [24] Johnston, A.C., Warkentin, M., McBride, M. & Carter, L. (2016), 'Dispositional and situational factors: influences on information security policy violations', *European Journal of Information Systems*, 25(3), pp. 231–251.
- [25] Kaspersky (2022), *Cybersicherheit in der Supply Chain Deutschlands - Aktuelle Kaspersky-Studie legt Status Quo der IT-Sicherheit in deutschen Unternehmen offen*. (https://go.kaspersky.com/supply-chain-report-de.html?utm_medium=PR. Accessed: 18.03.2023.
- [26] Kergel, D. (2018) *Qualitative Bildungsforschung: Ein integrativer Ansatz*. Wiesbaden: Springer Fachmedien Wiesbaden.
- [27] Kim, S.H. & Kwon, J. (2019), 'How Do EHRs and a Meaningful Use Initiative Affect Breaches of Patient Information?', *Information Systems Research*,
- [28] Knight, R. & Nurse, J.R.C. (2020), 'A framework for effective corporate communication after cyber security incidents', *Computers & Security*, 99pp. 1–18.
- [29] Kuckartz, U. & Rädiker, S. (2022), in *Qualitative Inhaltsanalyse: Methoden, Praxis, Computerunterstützung: Grundlagentexte Methoden*. Grundlagentexte Methoden. 5. Auflage. Weinheim Basel: Beltz Juventa.
- [30] Kude, T., Hoehle, H. & Sykes, T.A. (2017), 'Big data breaches and customer compensation strategies: Personality traits and social influence as antecedents of perceived compensation', *International Journal of Operations & Production Management*, 37(1), pp. 56–74.
- [31] Kulynych, J. & Korn, D. (2002), 'The Effect of the New Federal Medical-Privacy Rule on Research', *New England Journal of Medicine*, 346(3), pp. 201–204.
- [32] Kwon, J. & Johnson, M.E. (2015), 'Protecting Patient Data-The Economic Perspective of Healthcare Security', *IEEE Security & Privacy*, 13(5), pp. 90–95.
- [33] Lu, K. & Liao, H. (2023), 'Dynamic preference elicitation of customer behaviours in e-commerce from online reviews based on expectation confirmation theory', *Economic Research-Ekonomska Istraživanja*, 36(1), pp. 2915–2938.

- [34] Malhotra, A. & Kubowicz Malhotra, C. (2011), 'Evaluating Customer Information Breaches as Service Failures: An Event Study Approach', *Journal of Service Research*, 14(1), pp. 44–59.
- [35] Martin, K.D., Borah, A. & Palmatier, R.W. (2017), 'Data Privacy: Effects on Customer and Firm Performance', *Journal of Marketing*, 81(1), pp. 36–58.
- [36] Masuch, K., Greve, M. & Trang, S. (2021), 'What to do after a data breach? Examining apology and compensation as response strategies for health service providers', *Electronic Markets*, 31(4), pp. 829–848.
- [37] Odusote, A. (2021), 'Data Misuse, Data Theft and Data Protection in Nigeria: A Call for a More Robust and More Effective Legislation', *Beijing Law Review*, 12(04), pp. 1284–1298.
- [38] Oliver, R.L. (1980), 'A Cognitive Model of the Antecedents and Consequences of Satisfaction Decisions', *Journal of Marketing Research*, 17(4), pp. 460–469.
- [39] Ong, R. & Sabapathy, S. (2021), 'Hong Kong's data breach notification scheme: From the stakeholders' perspectives', *Computer Law & Security Review*, 42pp. 1–16.
- [40] Otto, P.N., Anton, A.I. & Baumer, D.L. (2007), 'The ChoicePoint Dilemma: How Data Brokers Should Handle the Privacy of Personal Information', *IEEE Security & Privacy Magazine*, 5(5), pp. 15–23.
- [41] Ponemon Institute (2022), Bericht über die Kosten einer Datenschutzverletzung 2022, IBM Security, <https://www.ibm.com/reports/data-breach>. Accessed: 18.03.2023.
- [42] Posey, C., Raja, U., Crossler, R.E. & Burns, A.J. (2017), 'Taking stock of organisations' protection of privacy: categorising and assessing threats to personally identifiable information in the USA', *European Journal of Information Systems*, 26(6), pp. 585–604.
- [43] Roberds, W. & Schreft, S.L. (2009), 'Data breaches and identity theft', *Journal of Monetary Economics*, 56(7), pp. 918–929.
- [44] Spreng, R.A., MacKenzie, S.B. & Olshavsky, R.W. (1996), 'A Reexamination of the Determinants of Consumer Satisfaction', *Journal of Marketing*, 60(3), pp. 15–32.
- [45] Strauss, A.L. & Corbin, J.M. (1998), *Basics of qualitative research: techniques and procedures for developing grounded theory*, in 2nd ed. Thousand Oaks: Sage Publications.
- [46] Strong, K.C., Ringer, R.C. & Taylor, S.A. (2001), 'THE* Rules of Stakeholder Satisfaction (* Timeliness, Honesty, Empathy)', *Journal of Business Ethics*, 32pp. 219–230.
- [47] Tanimura, J.K. & Wehrly, E.W. (2009), *The Market Value and Reputational Effects from Lost Confidential Information*.
- [48] Whitler, K.A. & Farris, P.W. (2017), 'The Impact of Cyber Attacks On Brand Image: Why Proactive Marketing Expertise Is Needed for Managing Data Breaches', *Journal of Advertising Research*, 57(1), pp. 3–9.
- [49] Widup, S., Rudis, B., Hylender, D., Spitler, M., Thompson, K., Baker, W., Bassett, G., Karambelkar, B., Brannon, S., Kennedy, D. & Jacobs, J. (2015), *2015 Verizon Data Breach Investigations Report*.
- [50] Witzel, A. (2000), 'Das problemzentrierte Interview', in *Qualitative Social Research. Sozialforschung*, p. 13.
- [51] Zhu, K. (2002), 'Information Transparency in Electronic Marketplaces: Why Data Transparency May Hinder the Adoption of B2B Exchanges', *Electronic Markets*, 12(2), pp. 92–99.
- [52] Zviran, M. & Haga, W.J. (1999), 'Password Security: An Empirical Study', *Journal of Management Information Systems*, 15(4), pp. 161–185.

Appendix

Table 4
Statements of customers' expectations of the company's response

Category	Respondent	Statement
Compensation	B3	"If something had really happened, that someone had debited money from my account, then yes. But if I don't have any other obvious damage, then I don't think they have to give me anything back. That would be nice, but I don't think it's that important."
	B5	"With the negligence one, there definitely, I do expect compensation at the end."
	B10	"I expect a high level of compensation if personal data is published that is problematic or has far-reaching consequences for my working life. That means, for example, if I explain my relationships with some colleagues or bosses or whatever, and this is published, and as a result I have restrictions within my job: I am somehow removed because they have read how I think about some things or things like that, then I definitely expect a high level of compensation. But if it just comes out where I'm going to spend the next summer vacation, the compensation can definitely be lower."
	B4	"If I had suffered any financial loss, I would have definitely expected something like [a compensation]"
	B3	"It is much more sensitive data, my expectations are then already higher and also compensation"
	B1	"There is the possibility of the health insurance to receive services that are beneficial to your health, such as in the form of fitness programs, cures or much more, which must be taken over by each health insurance patient or each patient with to some extent, which they could then perhaps provide free of charge."
Notification	B11	"[I] would also want to know how it came about now."
	B5	"If the investigation drags on, [I] expect to receive updates on its progress."
	B8	"I would like to know exactly what data was misused and to what extent."
Follow-up Notification	B10	"I would always like to be informed about the next steps and that we might be able to talk a bit about this."
Fast Reaction of the Company	B1	"Act as fast as possible, certainly, and contact affected parties and get it fixed as soon as possible."
	B4	"By the time it comes out through the media, it's actually too late."
	B6	"If more sensitive data is involved, I expect to be notified in a timely manner."
Transparency	B3	"The more sensitive the data becomes, the more important it is (...) that companies are transparent (...)"
	B5	"If it drags on for a long time, I definitely also expect to be told in between how it looks and definitely at the end the result, the conclusion (...)"
		"[Transparency] is simply much more important. That you really get the feeling that you are not so much in danger."

Table 4
Statements of customers' expectations of the company's response (continued)

Category	Respondent	Statement
Apology	B4	"If it is obvious that the company has nothing to do with it, then I don't expect an apology."
	B7	"I would actually also consider whether I change the health insurance company then."
	B5	"I expect on both points that they contact me, give me info on how they want to proceed and apologize that it happened. So that's what I expect in both cases, even though it's not their fault."
	B1	"A company can also apologize here only I think in writing personally to one."
Empathy	B7	"The more empathetic at this point, the better. If you come back to it in the mail or letter and say, we've tried our best, we've implemented the possible security standards, and the attacker still succeeded, I think I have a little more understanding."
	B7	"As long as (...) an empathic apology comes for it."
Measures	B4	"That they'll make sure it never happens again."
	B2	"A general information or such a preventive screening to (...) check the passwords that were used (...), check them against such a database of public passwords."
	B4	"There's also two-factor authentication or something like that, but Zalando doesn't have that, which makes it a bit more secure somehow (...)."
	B8	"That it is ensured that this will definitely no longer occur in the future, because this is already very sensitive data, where I would also like to see a higher level of data protection than is currently the case with Instagram, for example."
Support	B10	"What consequences, what I could have to fear, how you would advise me, how I should best proceed regarding my data breaches."
	B4	"That they may also give me advice on how I could improve the security."
	B8	"How one is informed preventively."
	B9	"Support from the company (...), how I can proceed further and what I can do against it, and what the possibilities are for me."
Participation in the Decision-making Process	B1	"I would like to find a way to satisfy both parties (...) [and] would like to participate in the decision-making process."
	B4	"In the decision-making process, they could provide me with more options (...) [such as] the option to receive a voucher, a special membership, or free shipping."
	B2	"[And offer] some flexibility and work with you to find a solution that suits you."

Table 5
Statements on the factors influencing customer expectations

Category	Respondent	Statement
Severity of the Data Breach	B1	"Depending on the severity, [I would] (...) switch my health insurance company."
	B10	"If it has far-reaching consequences, I would really consider ending my relationship with the company."
	B3	"In the difficult case [I want to] get even more information about what was done to restore everything. So, in this case, I consider all the information even more important."
Type of Company	B7	"Because this is already very sensitive data, where I would also like to see a higher level of data protection than now, for example, with Instagram. Exactly. So, my expectations of health insurance are significantly higher."
	B4	"Because of the size of the company, I would definitely expect them to deal with it in a transparent way, that they keep following up on the data breach."
	B4	"But if it's not such a large company (...) and if you compare it with a small or medium-sized company, then you might be a bit more understanding for such a data breach."
Type of Data	B11	"Such data is of course very sensitive data and especially in relation to future employers, etc. It is of course difficult when such data is used. Of course, it is difficult when such data is made public."
	B6	"Yes, if more sensitive data is affected, I expect fast notification."
	B11	"Yes, because it is health data with which you can do a lot. And as I said, for example, that the employer might not consider you, I think that's definitely data where you should take very strong precautions and protective measures so that it doesn't get out to the public."
	B2	"If I don't have any other obvious damage, then they don't have to give me anything back. It would be nice, but I don't think it's that important [and] (...) if money should really be withdrawn, then (...) [I also want] the same amount of money back."
	B3	"The password was probably stolen and leaked at some point, but there was no intrusion by any third party trying to gain access to the account. That's why I wouldn't expect anything more based on my experience."
Personal Responsibility of the Customer	B7	"And I also assume that it will be the case at some point. But since I myself am also to a certain extent to blame for the information and data that I share and that can also become public, and I also believe that Instagram itself can do little about it, I actually don't expect that much at all."
	B4	"If it's kind of your own fault that something like that happens, that you change passwords more often or email addresses or something."
Company Fault	B4	"Yes, so if I clicked on some phishing email and then my data was stolen, then of course I don't expect compensation from the company. (...) So if it's clear that the company has nothing to do with it, then I don't expect an apology."

Table 5
Statements on the factors influencing customer expectations (continued)

Category	Respondent	Statement
Company Fault	B5	“For example, I think I would prefer they already have a solution. I think that if there is a problem that was perhaps caused by them because they had a security gap, then I think they must also come up with the optimum solution for me afterwards. And that's where I tend not to want to be involved in the decision-making process because I don't want to be involved that much.”
	B5	“So definitely, I would have much higher expectations if the company was negligent. So much higher. If I expect (...) that they contact me, give me information on how they want to proceed and apologize that it happened. (...) and with the negligence, there in any case, I already expect compensation in the end.”
	B2	“So, I wouldn't expect it, I wouldn't take it for granted. I would definitely appreciate it. Especially if we don't necessarily take Coinbase as an example now, but any data breakdowns where it was really the fault of the company itself in the past, because someone screwed up.”

Table 6
Statements on meeting the customer expectation

Category	Respondent	Statement
Meeting the Customer Expectations	B6	“Because you're disappointed, simply. You had expectations, they are not fulfilled or just destroyed. And yes, then I am sad and that is normal.”
	B2	“Then you go to the competition. Simple as that,”
	B2	“That would confirm me or at least confirm my opinion that I made the decision to become a customer of this company at the time.”
	B8	“Then I would definitely be satisfied (...). My trust (...) would be greater than before. (...) [I] would not worry that something like this will happen again or look for alternatives.”

Integrating IT Security Aspects into Business Process Models: A Taxonomy of BPMN Extensions

Leonard Nake¹

¹ Martin Luther University Halle-Wittenberg, Universitaetsring 3, 06108 Halle (Saale), Germany

Abstract

Protecting themselves from IT security breaches is a crucial and cost-intensive task for the organizations of today. To achieve this, organizations implement bundles of IT security measures to secure their assets, which substantially influences their business processes. Therefore, aspects from the IT security domain must be integrated into business process models to adequately represent reality. There are various papers introducing extensions that integrate these aspects into the Business Process Model and Notation (BPMN) language. However, existing literature reviews are outdated and do not identify common characteristics among BPMN extensions that integrate IT security aspects. Based on the analysis of 18 papers that were identified during a structured literature review, this article develops a multi-dimensional taxonomy of BPMN extensions. This taxonomy identifies common characteristics and dimensions of the extensions and therefore gives a structured overview of the field and provides profound insights into the existing work.

Keywords

IT security, BPMN extension, literature review, taxonomy

1. Introduction


Technological innovations, such as intelligent process automation or cloud computing, have drastically changed the business processes of companies in the last years and provided them with opportunities to develop competitive advantages. However, these innovations also introduce new security risks that need to be addressed. Technologies and other organizational assets must be protected from attacks aiming to access sensitive information, change the data in information systems, and disrupt the normal operations of information systems [1]. It is typically not sufficient to implement isolated IT security measures for single assets as complex bundles of interdependent measures are required. Because of this, IT security measures have a substantial influence on the business processes of organizations [2]. Hence, aspects of this highly influential IT security domain that are addressed with such measures should be integrated into business process models for them to adequately represent reality.


Because of this need to integrate IT security aspects into business process models, many extensions for Business Process Model and Notation (BPMN), which is the de facto standard modeling language for business process models, have been introduced. While these extensions have similarities to one another, each approaches the problem from a different angle and therefore defines different concepts to integrate IT security aspects. For instance, there are approaches that extend BPMN with the necessary attributes to perform risk assessments [3, 4] while other approaches extend it with administrative control policies such as the separation of duty [5]. While there are papers that conduct literature reviews in this field [6, 7], they are outdated as many extensions have been published since their publications. Additionally, they do not identify common characteristics among the identified BPMN extensions. Therefore, this article aims to firstly identify the latest and relevant BPMN extensions in the field and secondly

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ leonard.nake@wiwi.uni-halle.de (L. Nake)

ORCID  0000-0001-8324-5641 (L. Nake)

 © 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

create a taxonomy to identify the characteristics and dimensions of the different extensions to give a structured analysis of the existing literature and show research gaps. I raise the following research question to address this problem:

What are the dimensions and characteristics of BPMN extensions that integrate IT security aspects into business process models?

To address the research question, I base the research design of the literature review on the method proposed by vom Brocke et al. [8]. The taxonomy creation is based on the method proposed by Nickerson et al. [9]. The contribution of this article is the new domain knowledge introduced through the literature review, on the one hand, and through the rigorous creation of the taxonomy, on the other hand. This paper is structured in the following way: In section 2, the related research relevant to this topic is discussed. The research design is explained in section 3 in detail. Section 4 describes the taxonomy that was created from the identified BPMN extensions. Section 5 is a further discussion of the findings of the literature review and the creation of the taxonomy. Section 6 concludes this article.

2. Related Research

The related research of my paper can be divided into two types. Considering the first type, there are two publications that conduct reviews of BPMN extensions that integrate IT security aspects into business process models in some way. In their research, Maines et al. [7] analyze BPMN-security extensions to create a cyber security ontology. Although this might seem quite similar to my research, it is quite different. Maines et al. [7] focus on extensions that provide BPMN-security instead of IT security in general. Therefore, the identified literature as well as the goal of the research is not the same. For instance, the identified literature that extends BPMN to conduct risk assessments [10, 11] are not considered in the ontology. Additionally, Maines et al. [7] create an ontology instead of a taxonomy and their research was done in 2015, which means that several newer BPMN extensions could not be considered in their research.

Other closely related research is the paper of Leitner et al. [6]. The authors conduct a literature review of security aspects in BPMN and provide an overview of the identified concepts in combination with the extended BPMN elements. This paper is the most similar to my research and the identified concepts are still relevant today, which is why it was analyzed during the first iteration of creating the taxonomy. However, the literature review was done in 2013. Since then, many new BPMN extensions have been published. Therefore, a new literature review was necessary to identify all relevant extensions. Additionally, my paper has the goal to create a taxonomy from the identified BPMN extensions by using the well-known method developed by Nickerson et al. [9] to identify common characteristics among the extensions.

The literature identified during my literature search is the second type of related research. It is analyzed in the following sections.

3. Research Method

3.1. Structured Literature Review

This study aims to create a taxonomy of existing BPMN extensions that integrate IT Security aspects into business process models and therefore needs to identify the relevant literature. To achieve this, I conducted an exhaustive but selective structured literature review [12] and followed the methodological guidelines of Webster & Watson [13] as well as vom Brocke et al. [8]. In their research, vom Brocke et al. [8] define five steps necessary for a structured literature review. As the first step the review scope has to be defined. For my review, I defined the scope as articles that introduce or discuss BPMN extensions that deal with IT Security in some way. These articles must be published between 2013 and 2023 as older papers were already identified in the

work of Leitner et al. [6], must be in English, and must be published in established scientific databases (ACM Digital Library, AIS Electronic Library, SpringerLink, IEEE Xplore, ScienceDirect). The second step is to conceptualize the topic. This work focuses on BPMN extensions since BPMN is the de facto standard language for modeling business processes. I also researched definitions and synonyms for IT Security. The third step is the actual literature search. The search string ("BPMN" AND ("IT Security" OR "information security") AND "extension") that was used to search full texts of articles resulted in 530 hits over the five databases. Then, the titles of the publications were analyzed which led to a drastic reduction in the number of hits (see Figure 1). After analyzing the abstracts and full text, there were 13 relevant articles left. We excluded articles that do not introduce or discuss BPMN extensions dealing with IT Security aspects but use other modeling languages or do not have IT Security as a main focus. As proposed by Webster & Watson [13], we then conducted a backward search which resulted in 7 additional papers after removing duplicates and analyzing the abstracts as well as the full texts. This led to a total number of 20 articles. Of these 20 publications, 18 introduced or improved relevant BPMN extensions and 2 articles reviewed the topic (see section 2). The fourth step is the analysis and synthesis of the identified literature. To achieve this, a taxonomy is developed in section four. Finally, in the last step, a research agenda has to be developed. This is done by discussing possibilities for future research in section six.

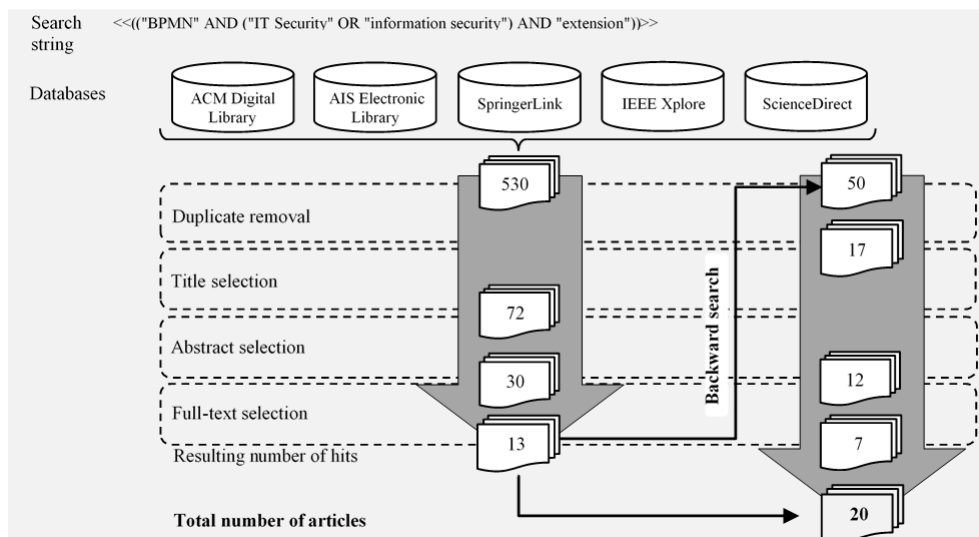


Figure 1: Literature Search Process

3.2. Development of Taxonomies in Information Systems

After identifying the relevant literature, I created a taxonomy following the widely used method for taxonomy development in information systems from Nickerson et al. [9]. The method consists of seven steps as shown in figure 2. The first step is to define a meta-characteristic that is based on the purpose, the users, and the expected use of the taxonomy. All characteristics must be logical consequences of this meta-characteristic. The second step is to determine the ending conditions for the taxonomy development. Then, one of two approaches has to be selected. The conceptual-to-empirical approach focuses on the conceptualization of dimensions and characteristics without examining the actual objects. This means that the creation of dimensions and characteristics is based on the researcher's notions about how the objects are similar and dissimilar. In the empirical-to-conceptual approach, a researcher has to identify a set of objects for the classification. Then, the researcher analyses these objects to find common characteristics and dimensions among them. Both approaches lead to the creation of a taxonomy that has to be evaluated considering the ending conditions. If all ending conditions are met the taxonomy development ends. If not all ending conditions are met, more conceptual-to-empirical or empirical-to-conceptual iterations have to be conducted.

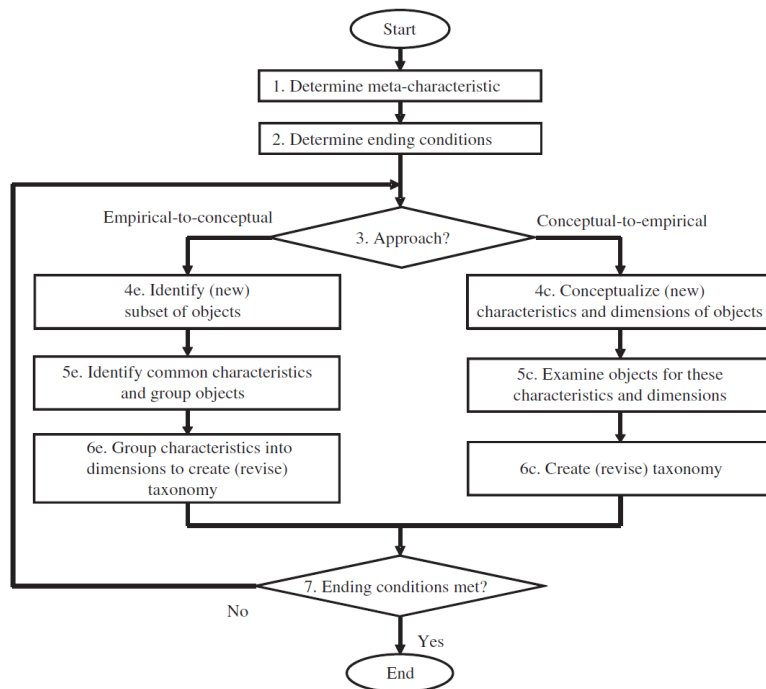


Figure 2: Taxonomy Development Method by Nickerson et al. [9]

Nickerson et al. [9] proposed 13 ending conditions that have to be met. There are eight objective ending conditions (all objects examined, no objects/dimensions/characteristics split or merged in the last iteration, at least one object under every characteristic, no new dimensions/characteristics in the last iteration, each dimension/characteristic is unique, and each combination of characteristics is unique) and five subjective ending conditions (concise, robust, comprehensive, extendible, and explanatory). I describe the development of my taxonomy in the next section.

4. Taxonomy of BPMN Extensions Integrating IT Security Aspects into Business Process Models

The creation of the taxonomy required 3 iterations until all ending conditions were met. The first conducted iteration was conceptual-to-empirical. In this iteration, I analyzed a literature review about security aspects in BPMN [6] to conceptualize dimensions and characteristics from the discussion of the literature in this article as a first step. Then, other theoretical work about the topic was consulted to gain insights into possible dimensions and characteristics. The other two iterations were empirical-to-conceptual. During these iterations, the 18 identified articles were analyzed to identify common characteristics. After the second iteration, three dimensions were discarded as they did not provide meaningful insights. Also, the dimension domain specificity had to be added since articles that introduced domain specific attributes differed significantly from more generic IT security attributes. While conducting the third iteration, all ending conditions were met. Neither dimensions nor characteristics changed during this iteration although all objects found during the literature search were classified. The resulting taxonomy is concise, robust, comprehensive, extendible, and explanatory and does not consist of repetitive characteristics or dimensions. It consists of five dimensions with 30 characteristics that are described in section 4.2. Of the five dimensions, only the domain specificity has mutually exclusive characteristics. This is a deliberate decision to make the taxonomy more concise and useful.

4.1. Meta-Characteristic

A meta-characteristic acts as the basis for the choice of characteristics so that each characteristic logically follows the previously defined meta-characteristic [9]. Since our taxonomy is aimed at researchers and practitioners that want to integrate IT security aspects into their business processes, we want to give a practical overview instead of going too far into the technical details of each extension. Hence, I define the meta-characteristic of the taxonomy as: 'characteristics of IT Security aspects and their extended BPMN elements defined in the identified BPMN extensions from a functionality perspective'.

4.2. Dimensions and Characteristics

The first dimension **risk assessment** includes articles that extend BPMN by aspects needed for conducting security risk assessments [4, 10, 11]. The characteristics are the security aspects used to perform these risk assessments. Reliability in this context is the counterpart to the failure probability. Papers extend BPMN with this value to include the probability of a security incident [3]. Risk objective describes the maximum value of acceptable risk in a business process [11]. Risk information is the risk value of a process or task based on the values of reliability and asset value [3]. Vulnerabilities as a characteristic means that the paper extends BPMN with information about vulnerabilities of a business process, for example, an insecure communication protocol [11]. The Asset value corresponds to the value that an asset represents for the organization [3]. While these characteristics are all part of the risk assessment, it makes sense to include them as separate characteristics in the taxonomy because the papers differ in the way they perform the assessment. Additionally, other BPMN extensions can integrate only some of the aspects into the business processes. For instance, Altuhhov et al. [14] introduced an annotation called "vulnerability point" to mark vulnerable assets, such as data objects or tasks.

The dimension **task execution rules** is comprised of rules about the execution of tasks. Separation of duty means that a task cannot be executed by a single person but has to be executed by at least two persons. The binding of duty dictates that several tasks have to be executed by the same person [15, 16]. The third characteristic is the rule non-delegation which means that a task can only be executed by assigned users [17].

The dimension **security goal** is built on the RMIAS reference model developed in the work of Cherdantseva & Hilton [18] and is referenced frequently in the different articles. It involves the following characteristics:

Authenticity describes the ability of a system to verify identity and establish trust in a third party as well as in the provided information [18]. The analyzed BPMN extensions try to implement this principle in different ways. For instance, Salnitri et al. [19] impose that the identity or authenticity of a user has to be verified in activities by requiring executors to have a minimum level of trust or by banning anonymous users from executing activities. Authenticity is also defined for data objects. Using the extension makes it possible to prove the genuineness of the data object by proving that the data was not modified by unauthorized parties or by proving the identity of the entity who generated or modified it. Salnitri et al. [19] give the example of a visa as a data object that is marked with an authenticity annotation that specifies the security mechanisms TLS (Transport Layer Security) and X.509 to be used in order to guarantee the integrity of the visa data.

Availability means that a system needs to ensure that all its components are available and functional when they are required [18]. One instantiation of availability found in the literature tries to ensure that critical resources are always available to process participants. If a requested resource is not available the system has to maintain backups from which the respective data object can be retrieved so that it is always available for the user [20].

Accountability describes a system's ability to hold users accountable should they perform harmful actions [18]. One of the ways how accountability is achieved in business process models is described in the work of Argyropoulos et al. [20]. In their extension, only process participants

with appropriate permissions can access resources or perform certain activities if they are authorization constrained.

Auditability means that a system needs to monitor all actions performed by actors in the system in a way that it cannot be bypassed [18]. There are different ways to implement auditability in business process models. For activities, it can be made possible to save all the actions performed by the executor of an activity. For data objects, it can be made possible to keep track of all actions concerning the data object, such as write, read, or store. For a message flow, it can be made possible to save all the actions performed during the communication [21].

Confidentiality is a system's ability to make information only accessible to authorized users [18]. One way to guarantee confidentiality is introduced by Pullonen et al. [22]. Their extension allows for the encryption and decryption of data in so-called privacy-enhancing technology tasks. It uses a data input and a public key to generate a ciphertext that can be decrypted with the respective secret key.

Integrity describes the ability of a system to ensure completeness, accuracy as well as the absence of unauthorized modifications in its components [18]. One example of an implementation of integrity in a business process model using a BPMN extension is to compare the system's copy of data to the original by data validation techniques if the data object in the business process model is integrity-constrained [20].

Non-repudiation means that a system needs to have the ability to prove the occurrence or non-occurrence of events and the participation or non-participation of parties in this event [18]. An example of non-repudiation in a business process model is described by Salnitri et al. [21]. For activities, the execution and non-execution of an activity can be made provable. For message flows, it can be made verifiable if a message flow was used or not used.

Privacy is a system's duty to obey privacy legislation. The system needs to enable individuals to control their personal information if feasible [18]. Privacy can be introduced to business process models by specifying that activities or data objects must be compliant with privacy legislation and should therefore let users control their own data [19].

The dimension **domain specificity** has the characteristics generic and domain specific and it describes whether the BPMN extension contains attributes that do not only implement generic IT security aspects but also domain specific aspects. Most BPMN extensions introduce generic concepts that exclusively implement IT security aspects into the business process model. However, there are exceptions in the identified literature. For instance, in addition to similar IT security aspects Ramadan et al. [17] introduce annotations for anonymity, undetectability, unlinkability, unobservability, and fairness for including data-minimization and fairness in the business process model. Köpke et al. [22] introduce annotations for enforceability and privacy in their model-driven approach to designing secure smart contracts.

The dimension **extended BPMN element** describes which of the existing BPMN elements were extended by each extension. It consists of the characteristics Activity, Event, Gateway, Pool, Message Flow, Data Object, Process, Subprocess, and Other that refer to the elements defined in the BPMN language. All extensions extended activities in some way and many extended data objects. Other BPMN elements were extended more rarely. The decision about which element is extended depends on the goal of each extension. For instance, Varela-Vaca et al. [4, 10, 11] decided to integrate most of their parameters by extending pools since they see the business process inside a pool as the main asset that needs assessment for their approach.

Risk Assessment	Reliability (5)	Risk Objective (5)	Risk Information (5)
	Vulnerabilities (4)	Asset Value (5)	None (12)
Task Execution Rules	Separation of Duty (5)		Binding of Duty (5)
	Non-delegation (3)		None (13)
Security Goal	Accountability (8)	Auditability (6)	Authenticity (10)
	Confidentiality (10)	Integrity (11)	Availability (11)
	Non-repudiation (7)	Privacy (8)	None (5)
Domain Specificity	Domain Specific (3)		Generic (15)
Extended BPMN Element	Activity (18)	Event (4)	Gateway (3)
	Pool (6)	Message Flow (7)	Data Object (12)
	Process (2)	Subprocess (2)	Other (6)

Table 1: Taxonomy of BPMN Extensions for Integrating IT Security Aspects

5. Discussion

Table 1 shows the final taxonomy. The small numbers in brackets show how many of the BPMN extensions fulfill each characteristic. If an extension integrates the respective IT security aspect it is counted into this number. The classification of the papers shows that most BPMN extensions are generic, meaning that they do not introduce domain specific but general IT security aspects. Most extensions perform no risk assessment but introduce annotations and execution logic into the business process model to achieve the security goals defined in [18]. Some extensions implement task execution roles. For example, some require tasks to be executed by at least two persons. While the BPMN elements that are extended differ in the different articles, all of them extend activities in some way.

There is a clear distinction in the identified literature between risk-oriented BPMN extensions and security goal-oriented extensions. The former focus on implementing security risk-related data into the business process model to perform calculations for a risk assessment. The latter focus on annotating and regulating BPMN elements to achieve security goals during the execution of the business process itself. The risk-oriented extensions are defined by Varela-Vaca et al. [4, 10, 11] and Cardoso et al. [3, 23]. While they aim for similar goals, there are differences. Varela-Vaca et al. [4, 10, 11] extend BPMN to assess the conformance of IT security properties in business process models by adding new calculations and model logic. Cardoso et al. [3, 23] focus on extending BPMN using the standard to perform quantitative risk assessment. Among the security goal-oriented extensions the most used language is SecBPMN [19]. On the one hand, it was refined by the authors themselves [5, 21]. On the other hand, several other publications referenced SecBPMN and augmented it with other aspects [17, 22]. Extensions built on the SecBPMN language are the only objects that fulfill all characteristics of the security goal as well as the task execution rules dimension. Therefore, it seems to be the standard extension in the field. Still, many authors published their own extensions to address the specific problems of their research fields [24–27]. The taxonomy shows that there seem to be research gaps in the field. The most

obvious one is that the risk-oriented extensions barely apply security goal-oriented concepts and vice versa. Naturally, it is possible to implement extensions from both groups at the same time but this would mean that the concepts do not influence each other. In reality, changes in the business process model caused by security goal-oriented concepts could influence the risk of the underlying business process. To integrate IT security aspects holistically it could be beneficial to consider both orientations. Therefore, it could be an interesting research objective to combine the two groups in a new extension.

6. Conclusion

In this research, I conducted a rigorous, structured literature review that led to the identification of 18 papers introducing BPMN extensions that integrate IT security aspects into business process models and two papers that review such extensions. Then, I created a multidimensional taxonomy from these 18 papers to answer my research question. I derived five dimensions and 30 characteristics of BPMN extensions that integrate IT security aspects that are explained in detail in section 4.2. However, there are limitations to this research that need to be considered. I followed well-known methods for the literature review as well as for the taxonomy creation to guarantee scientific rigor and maximize the objectivity of the research. Nevertheless, the results of this research are influenced by subjective decisions. Firstly, the literature review was done in only five scientific databases and the exclusion of papers is subjective to a certain extent. Additionally, the selection of the search string used in the literature search is partly subjective. However, I experimented with different synonyms (for example, expansion and augmentation as synonyms for extension) to analyze and improve the search results. Secondly, the actual creation of the taxonomy with all its dimensions and characteristics is a subjective process. Therefore, it is possible that other researchers would have developed other characteristics and dimensions. Despite these limitations, I believe that this work provides useful insights for scientists and practitioners.

This paper makes the following scientific contributions. The developed taxonomy can be used as a basis for further research about BPMN extensions that integrate IT security aspects. It provides knowledge about the relevant literature and can be used to classify new extensions. Additionally, the taxonomy structures the research field by deriving common characteristics and dimensions and shows research gaps, such as the observation that there are no extensions that combine the risk-oriented and the security goal-oriented view, which could be necessary for integrating a holistic combination of IT security aspects into the business processes of an organization. Therefore, it is possible to derive new BPMN extensions from the taxonomy. This research also has implications for practice. It allows practitioners to get an overview of existing BPMN extensions and their implemented IT security aspects and therefore provides them with insights that can help when choosing an extension that addresses their respective needs.

Acknowledgements

The project on which this study is based was funded by the German Federal Ministry of Education and Research under grant number 16KIS1331. The responsibility for the content of this publication lies with the author.

References

- [1] M. Z. Gunduz, and R. Das, "Cyber-security on smart grid: Threats and potential solutions," *Computer Networks*, vol. 169, p. 107094, 2020, doi: 10.1016/j.comnet.2019.107094.
- [2] S. Kühnel, S. Sackmann, S. Trang, I. Nastjuk, T. Matschak, L. Niedzela, and L. Nake, "Towards a Business Process-based Economic Evaluation and Selection of IT Security Measures," CEUR Workshop Proceedings 2966, CEUR-WS.org 2021, pp. 7-21.

- [3] P. Cardoso, A. Respício, and D. Domingos, “riskaBPMN - a BPMN extension for risk assessment,” *Procedia Computer Science*, vol. 181, pp. 1247–1254, 2021, doi: 10.1016/j.procs.2021.01.324.
- [4] Á. J. Varela-Vaca, L. Parody, R. M. Gasca, and M. T. Gómez-López, “Automatic Verification and Diagnosis of Security Risk Assessments in Business Process Models,” *IEEE Access*, vol. 7, pp. 26448–26465, 2019, doi: 10.1109/ACCESS.2019.2901408.
- [5] M. Salnitri, E. Paja, and P. Giorgini, “Maintaining Secure Business Processes in Light of Socio-Technical Systems' Evolution,” in *2016 IEEE 24th International Requirements Engineering Conference workshops: Proceedings : 12-16 September 2016, Beijing, China*, Beijing, China, 2016, pp. 155–164.
- [6] M. Leitner, M. Miller, and S. Rinderle-Ma, “An Analysis and Evaluation of Security Aspects in the Business Process Model and Notation,” in *2013 International Conference on Availability, Reliability and Security*, 2013.
- [7] C. L. Maines, D. Llewellyn-Jones, S. Tang, and B. Zhou, “A Cyber Security Ontology for BPMN-Security Extensions,” in *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, 2015, pp. 1756–1763.
- [8] J. vom Brocke *et al.*, *Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process*. [Online]. Available: <https://www.alexandria.unisg.ch/handle/20.500.14171/75942>
- [9] R. C. Nickerson, U. Varshney, and J. Muntermann, “A method for taxonomy development and its application in information systems,” *European Journal of Information Systems*, vol. 22, no. 3, pp. 336–359, 2013, doi: 10.1057/ejis.2012.26.
- [10] A. J. Varela-Vaca, R. M. Gasca, and A. Jimenez-Ramirez, “A Model-Driven engineering approach with diagnosis of non-conformance of security objectives in business process models,” in *2011 Fifth International Conference on Research Challenges in Information Science (RCIS 2011): Gosier, [Guadeloupe], France, 19 - 21 May 2011 ; [proceedings, Gosier, France, 2011, pp. 1–6.*
- [11] A. J. Varela-Vaca, R. M. Gasca, and S. Pozo, “OPBUS: Risk-aware framework for the conformance of security-quality requirements in business processes,” in *Proceedings of the International Conference on Security and Cryptography*, 2011, pp. 370–374.
- [12] H. M. Cooper, “Organizing knowledge syntheses: A taxonomy of literature reviews,” (in En;en), *Knowledge in Society*, vol. 1, no. 1, pp. 104–126, 1988, doi: 10.1007/BF03177550.
- [13] J. Webster, and R. T. Watson, “Analyzing the Past to Prepare for the Future: Writing a Literature Review,” *MIS Quarterly*, vol. 26, no. 2, pp. xiii–xxiii, 2002. [Online]. Available: <http://www.jstor.org/stable/4132319>
- [14] O. Altuhhov, R. Matulevičius, and N. Ahmed, “An Extension of Business Process Model and Notation for Security Risk Management,” *International Journal of Information System Modeling and Design (IJISMD)*, vol. 4, no. 4, pp. 93–113, 2013. [Online]. Available: <https://ideas.repec.org/a/igg/jismd0/v4y2013i4p93-113.html>
- [15] A. D. Brucker, I. Hang, G. Lückemeyer, and R. Ruparel, “SecureBPMN,” in *Proceedings of the 17th ACM symposium on Access Control Models and Technologies*, New York, NY, USA, 2012.
- [16] W. Labda, N. Mehandjiev, and P. Sampaio, “Modeling of privacy-aware business processes in BPMN to protect personal data,” in *Proceedings of the 29th Annual ACM Symposium on Applied Computing*, Gyeongju Republic of Korea, 2014, pp. 1399–1405.
- [17] Q. Ramadan, D. Strüber, M. Salnitri, J. Jürjens, V. Riediger, and S. Staab, “A semi-automated BPMN-based framework for detecting conflicts between security, data-minimization, and fairness requirements,” *Softw Syst Model*, vol. 19, no. 5, pp. 1191–1227, 2020, doi: 10.1007/s10270-020-00781-x.
- [18] Y. Cherdantseva and J. Hilton, “A Reference Model of Information Assurance & Security,” in *2013 Eighth International Conference on Availability, Reliability and Security (ARES 2013): Regensburg, Germany, 2 - 9 [i.e. 2 - 6] September [2013 ; proceedings ; including workshops*, Regensburg, Germany, 2013, pp. 546–555.

- [19] M. Salnitri, F. Dalpiaz, and P. Giorgini, "Modeling and Verifying Security Policies in Business Processes," in *Lecture Notes in Business Information Processing*, vol. 175, *Enterprise, business-process and information systems modeling: 15th International Conference, BPMDS 2014, 19th International Conference, EMMSAD 2014, held at CAiSE 2014, Thessaloniki, Greece, June 16-17, 2014 ; Proceedings*, I. Bider, Ed., Heidelberg: Springer, 2014, pp. 200–214.
- [20] N. Argyropoulos, H. Mouratidis, and A. Fish, "Enhancing secure business process design with security process patterns," (in En;en), *Softw Syst Model*, vol. 19, no. 3, pp. 555–577, 2020, doi: 10.1007/s10270-019-00743-y.
- [21] M. Salnitri, F. Dalpiaz, and P. Giorgini, "Designing secure business processes with SecBPMN," *Softw Syst Model*, vol. 16, no. 3, pp. 737–757, 2017, doi: 10.1007/s10270-015-0499-4.
- [22] J. Köpke, G. Meroni, and M. Salnitri, "Designing secure business processes for blockchains with SecBPMN2BC," *Future Generation Computer Systems*, vol. 141, pp. 382–398, 2023, doi: 10.1016/j.future.2022.11.013.
- [23] P. B. Cardoso, D. Domingos, and A. Respício, "Contributions for risk assessment of IoT-aware business processes at different granularity levels," *Procedia Computer Science*, vol. 192, pp. 991–1000, 2021, doi: 10.1016/j.procs.2021.08.102.
- [24] M. E. A. Chergui and S. M. Benslimane, "A Valid BPMN Extension for Supporting Security Requirements Based on Cyber Security Ontology," in *Lecture Notes in Computer Science*, vol. 11163, *Model and data engineering: 8th International Conference, MEDI 2018, Marrakesh, Morocco, October 24–26, 2018 : proceedings*, E. H. Abdelwahed, L. Bellatreche, M. Golfarelli, D. Méry, and C. Ordonez, Eds., Cham: Springer, 2018, pp. 219–232.
- [25] P. Pullonen, R. Matulevičius, and D. Bogdanov, "PE-BPMN: Privacy-Enhanced Business Process Model and Notation," in *Lecture Notes in Computer Science*, vol. 10445, *Business Process Management: 15th international Conference, BPM 2017 Barcelona, Spain, September 10-15, 2017 ; proceedings*, J. Llinás, G. Engels, and A. Kumar, Eds., Cham: Springer, 2017, pp. 40–56.
- [26] P. Pullonen, J. Tom, R. Matulevičius, and A. Toots, "Privacy-enhanced BPMN: enabling data privacy analysis in business processes models," *Softw Syst Model*, vol. 18, no. 6, pp. 3235–3264, 2019, doi: 10.1007/s10270-019-00718-z.
- [27] K. S. Sang, and B. Zhou, "BPMN Security Extensions for Healthcare Process," in *The 15th IEEE International Conference on Computer and Information Technology (CIT 2015), the 14th IEEE International Conference on Ubiquitous Computing and Communications (IUCC 2015), the 13th IEEE International Conference on Dependable, Autonomic and Secure Computing (DASC 2015), the 13th IEEE International Conference on Pervasive Intelligence and Computing (PICom 2015): CIT/IUCC/DASC/PICom 2015 : proceedings : 26-28 October 2015, Liverpool, United Kingdom, LIVERPOOL, United Kingdom, 2015*, pp. 2340–2345.

Appendix

Object	Risk Assessment					Task Execution Rules				Security Goal								Domain Specificity		Extended BPMN Element													
	Reliability	Risk Objective	Risk Information	Vulnerabilities	Asset Value	None	Separation of Duty	Binding of Duty	Non-delegation	None	Accountability	Auditability	Authenticity	Availability	Confidentiality	Integrity	Non-repudiation	Privacy	None	Domain Specificity	Generic	Activity	Event	Gateway	Pool	Message Flow	Data Object	Process	Subprocess	Other			
Brucker et al. 2012																																	
Varela-Vaca et al. 2019	x	x	x	x	x	x	x	x		x											x	x											
Varela-Vaca et al. 2011a	x	x	x	x	x																x	x											
Varela-Vaca et al. 2011b	x	x	x	x	x																x	x											
Cardoso et al. 2021a	x	x	x	x	x																x	x											
Cardoso et al. 2021b	x	x	x	x	x																x	x											
Saintri et al. 2014						x					x	x	x	x	x	x	x	x	x			x	x										
Saintri et al. 2016						x	x				x	x	x	x	x	x	x	x	x			x	x										
Pullonen et al. 2017						x					x	x	x	x	x	x	x	x	x			x	x										
Pullonen et al. 2019						x					x	x	x	x	x	x	x	x	x			x	x										
Argyropoulos et al. 2020						x					x	x	x	x	x	x	x	x	x			x	x										
Saintri et al. 2017						x					x	x	x	x	x	x	x	x	x			x	x										
Ramadan et al. 2020						x					x	x	x	x	x	x	x	x	x			x	x										
Altuhov et al. 2013											x	x	x	x	x	x	x	x	x			x	x										
Labda et al. 2014											x	x	x	x	x	x	x	x	x			x	x										
Sang & Zhou 2015											x	x	x	x	x	x	x	x	x			x	x										
Chergui et al. 2018											x	x	x	x	x	x	x	x	x			x	x										
Köpke et al. 2023											x	x	x	x	x	x	x	x	x			x	x										

From Pixels to Generalization: Ensuring Information Security and Model Performance with Design Principles for Synthetic Image Data in Deep Learning

Martin Böhmer¹

¹ Martin Luther University Halle-Wittenberg, Universitätsring 3, 06108 Halle (Saale), Germany

Abstract

This paper explores the ethical and effective utilization of synthetic image data in computer vision deep learning. It addresses the challenges of acquiring real-world training data and proposes design principles for selecting, generating, and integrating synthetic images. These principles cover aspects such as ethical compliance, privacy protection, scene diversity, and complexity management. By adopting a design science research approach and using a multi-method research design, the study provides actionable guidance for researchers and practitioners, as these design principles ensure responsible use of synthetic image data while improving model performance and privacy protection. The paper contributes to design knowledge in the general IS, deep learning, and IS ethics field, highlighting the theoretical and practical relevance of the proposed principles. The reusability of the design principles promotes the efficient use of synthetic image data in computer vision and has been positively evaluated.

Keywords

Design Principles, Deep Learning, Synthetic Image Data, Information Security, AI Ethics

1. Introduction

Since computer vision deep learning models often consist of millions or even billions of parameters, they rely on large amounts of training data to achieve high performance and generalization [1]. However, acquiring real-world training data for artificial intelligence (AI) applications can be costly, error-prone, limited, or imbalanced [2, 3, 4]. Synthetic image data (e.g. in the form of video game engine generated scenes) has emerged as a promising alternative, offering scalability, precision, and potentially more robust and accurate models [5, 2, 6]. Nonetheless, guiding design knowledge on how to utilize synthetically generated image data in deep learning remains scarce. Moreover, the synthetic illustration of humans, including their separate or related characteristics such as body parts, raises ethical considerations regarding privacy, consent, and the potential for misrepresentation or discrimination. In addition, the use-case context of synthetic image data often revolves around human-related domains [7, 8, 9], such as medicine or surveillance. These domains inherently involve sensitive information and human interactions, making it crucial to design technologies that align with ethical standards and user values. Given the nascent state of the synthetic imagery domain, this paper therefore defines the following guiding research question:


RQ: *How to ethically, effectively, and robustly utilize synthetic image data in computer vision deep learning environments?*

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ martin.boehmer@wiwi.uni-halle.com (M. Böhmer)

ORCID [0000-0002-6249-3184](https://orcid.org/0000-0002-6249-3184) (M. Böhmer)

© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

To answer this question and to contribute prescriptive knowledge, the design science research paradigm [10] and design science process model [11] are adopted, with a focus on the *value sensitive design* theory [12] as the guiding theoretical lens. The research approach employs a multi-method research design, combining qualitative methods such as moderated focus groups and think aloud sessions to ensure the validity and comprehensiveness of the study design. Therefore, the research aims to fill the aforementioned research gap by deriving design principles based on kernel theories, the literature, and practical insights. The design principles address key aspects such as ethical compliance, privacy protection, data governance, scene diversity, controlled composition, complexity management, and data augmentation, providing actionable guidance for researchers and practitioners in the selection, generation, and integration of synthetic image data for training deep learning models. By adopting these design principles, practitioners can ensure ethical and responsible use of synthetic image data while enhancing model performance, privacy protection, and generalization. The reusability of these principles in similar contexts contributes to their wider application and adoption, addressing the current lack of design knowledge and promoting efficient utilization of synthetic image data in computer vision deep learning environments.

The following sections present the research design, theoretical and practical foundations, design principles, and the evaluation of the proposed principles. The conclusion highlights the contributions of this study to the design knowledge in the field of utilizing synthetic image data in computer vision and identifies potential areas for future research.

2. Research Design

In order to contribute prescriptive rather than descriptive knowledge [13], the design science research paradigm [13] was used for the purpose of this study. In addition, value sensitive design theory [12] was used as the guiding theoretical lens (see Section 3.1), which served as the theoretical framework for the methodological techniques and design of the approach undertaken in this paper. Therefore, a multi-method research approach [14] consisting of different qualitative methods was chosen to address the shortcomings of single methods and to ensure the validity of the design. The methods used for this study are based on value sensitive design [12] and include a moderated focus group for data collection and a think aloud session for evaluating the design principles.

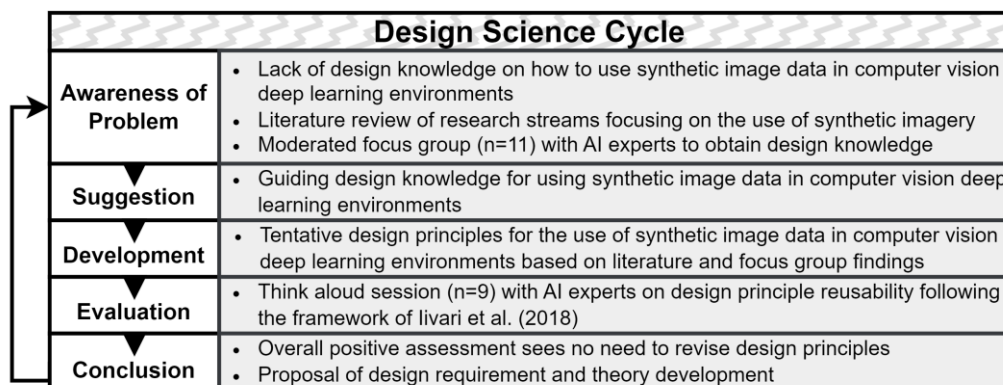


Figure 1: Design science research approach

Since the design science research paradigm can be operationalized through various methodological approaches [15], this study utilized the framework proposed by Vaishnavi and Kuechler [11] due to its explicit focus on theoretically grounded design principles. As shown in Figure 1, the approach includes the five steps: awareness of problem, suggestion, development, evaluation, and conclusion. Furthermore, and particularly with respect to the ethical and information security scope of this paper, the ethical design science framework proposed by Durani et al. [16] was used to derive ethical and, in addition, negative design principles (discussed

in detail in Section 4), addressing the disruptive nature of recent advancements in technology and especially deep learning. In this paper, we delve into the intricacies of design principles, because they form the foundation of design knowledge and play a pivotal role in solving the problem at hand [17], which in this case is the theoretical void of guiding design knowledge on how to use synthetic image data in deep learning.

3. Theoretical and Practical Foundations

Given that state-of-the-art deep learning models for computer vision comprise millions, if not billions, of parameters, the training process for these models necessitates an immense quantity of training data, which is frequently absent or imbalanced [1]. Therefore, the impulse for the underlying research stems from a genuine real-world circumstance, namely, the provision of scalable, precise, and ethical computer vision deep learning models and their respective training data. The highlighted problem originates from several prior studies that found synthetically trained computer vision models to be more robust, accurate, and less error-prone [5, 2, 6]. In addition, synthetic image data achieves photorealism and can be generated and scaled infinitely, making it a genuine alternative to conventional real imagery approaches [3]. Therefore, a thorough analysis of the scientific literature was conducted in the scope of this study to identify relevant research streams and kernel theories. In addition, the theoretical foundations are further supported by the results of a moderated focus group, which serve as practical foundations for the development of the design principles.

3.1. Kernel Theory

To ensure scientific rigor and stringency, design science research endeavors can use kernel theories to derive design principles. Broadly speaking, kernel theory functions as a form of justificatory knowledge within the realm of design knowledge development, as indicated by the work of Gregor and Hevner [10], such as in the form of design principles [18]. Henceforth, this study adopts *the analyze with lens*-mechanism proposed by Möller et al. [19], drawing upon the theoretical foundations of employing kernel theories as a means of analysis. The use of a theoretical lens allows researchers to derive concepts indirectly, guiding the analysis or framing of data within the conceptual borders of a specific theory. This approach aligns with the perspective of Niederman and March [20] on the theoretical lens, which emphasizes its role in aiding the theorization process, leading to the formulation of design principles or meta-requirements based on a data foundation. Thus, by adopting the *analyze with lens*-mechanism proposed by Möller et al. [19], this study aims to analyze the data through a theoretical lens, allowing for a more robust and informed exploration of the underlying concepts and patterns. As the most appropriate kernel theory for the scope of this study, the *value sensitive design* theory [12] was chosen as it epitomizes a theoretically grounded approach that considers human values in a principled and comprehensive manner throughout the design process, which aligns perfectly with the research goal of developing technology that respects and incorporates user values while ensuring ethically responsible and user-centered design decisions. In the specific use-case of synthetic image data utilization for deep learning tasks, this especially connects to the synthetic illustration of humans (including separate or related characteristics, e.g. body parts), the use-case context in which the synthetically generated image data is used (often human-related, e.g. medicine or surveillance), and the potential ethical implications that arise from the creation and utilization of synthetic images. The theoretical lens of *value sensitive design* [12] thus helps not only to derive design principles, but also to establish a robust elicitation and evaluation (e.g., think aloud sessions) of these principles. Moreover, the theory recognizes that technology is not neutral as it can influence behavior, perception, and societal structures, and thus should be designed in a way that reflects positive human values and respects ethical principles, promoting a holistic approach to technology design that goes beyond functionality in order to consider the broader impact on individuals, society, and the environment.

3.2. Theoretical Foundations

To establish the objectives for addressing the aforementioned problem, a comprehensive analysis of the scientific literature focused on the research area of synthetic image data generation in deep learning, in line with our DSR methodology, was conducted. As aforementioned, training deep learning models in computer vision requires large amounts of data to achieve a fairly high degree of generalization, which is often costly, missing, or unbalanced [1]. Several studies have highlighted the effectiveness of synthetic data for training deep learning models in various computer vision tasks. Lee et al. [21] and Krump et al. [6] utilized synthetic datasets for deep learning-based object detection, specifically in underwater sonar imaging and vehicle detection on UAV platforms, respectively. Body et al. [22] and Condrea et al. [23] demonstrated the value of artificial augmented textual data and purely synthetic training data, respectively, for sentiment analysis models and vital signs detection in videos. Similarly, Liu et al. [8] and Zaki et al. [24] employed synthetic data for pose estimation and semantic object/scene categorization. Hence, these studies collectively highlight the effectiveness of synthetic data in various computer vision domains.

Additionally, domain adaptation, which is a technique that involves adapting a model trained on one domain (synthetic) of data to perform well on a different but related domain (real), and transfer learning have been extensively explored in the context of synthetic data. Lahiri et al. [25], Venkateswara et al. [26], and Kuhnke and Ostermann [7] focused on unsupervised domain adaptation for synthetic data, learning transferable feature representations, and domain adaptation for pose estimation, respectively. Seib et al. [3] conducted a comprehensive review of current approaches that combine real and synthetic data to enhance neural network training, supporting the argument for a combination of training data and data augmentation. Aranjuelo et al. [27] discussed key strategies for synthetic data generation in people detection from omnidirectional cameras, emphasizing the effective use of both real and synthetic data. Valtchev and Wu [28] demonstrated the utility of domain randomization for neural network classification, showcasing the effectiveness of synthetic data in training robust models. These studies provide insights into the adaptation of synthetic data to real-world scenarios.

Moreover, the combination of synthetic and real training data has been investigated by several researchers. Wan et al. [29], Bird et al. [5], and Abu Alhaija et al. [30] utilized mixed datasets, comprising both synthetic and real data, for document layout analysis, scene classification, and object detection in augmented reality, respectively. Thereby, these studies highlight the benefits of leveraging both synthetic and real data for training computer vision models.

Furthermore, the use of synthetic data generation techniques and simulators has been explored. Müller et al. [31] introduced a photorealistic simulator for generating synthetic data for computer vision applications, whereas Zhang et al. [4] proposed a stacked multichannel autoencoder framework for efficient learning from synthetic data. Valerio Giuffrida et al. [32] generated synthetic training data for the detection of synthetic Arabidopsis plants using generative adversarial networks. Scheck et al. [33] introduced a synthetic dataset that serves as a valuable resource for training and evaluating deep learning models. These works provide insights into the generation and utilization of synthetic data for training deep learning models.

Despite the considerable research on utilizing synthetic image data for computer vision deep learning models, there remains a notable research gap in terms of a comprehensive framework or guidelines that provide design knowledge to effectively and systematically utilize synthetic data in this context. While individual studies have demonstrated the benefits and effectiveness of synthetic data in specific tasks, there is a lack of unified principles or guidelines that guide researchers and practitioners in the selection, generation, and integration of synthetic image data for training deep learning models. Hereby, the absence of such design knowledge hinders the widespread adoption and consistent utilization of synthetic data, leading to potential inefficiencies, suboptimal performance, and challenges in real-world deployment.

3.3. Practical Foundations

To ensure scientific rigor, and after analyzing the aforementioned research streams and kernel theory, it seemed reasonable to conduct a moderated focus group with AI experts to rigorously derive design knowledge, compare it to the literature findings, and incorporate it into the design principles. Moderated focus groups are especially predestined for extensive qualitative insights into a subject [34] and align with the kernel theory of value sensitive design [12].

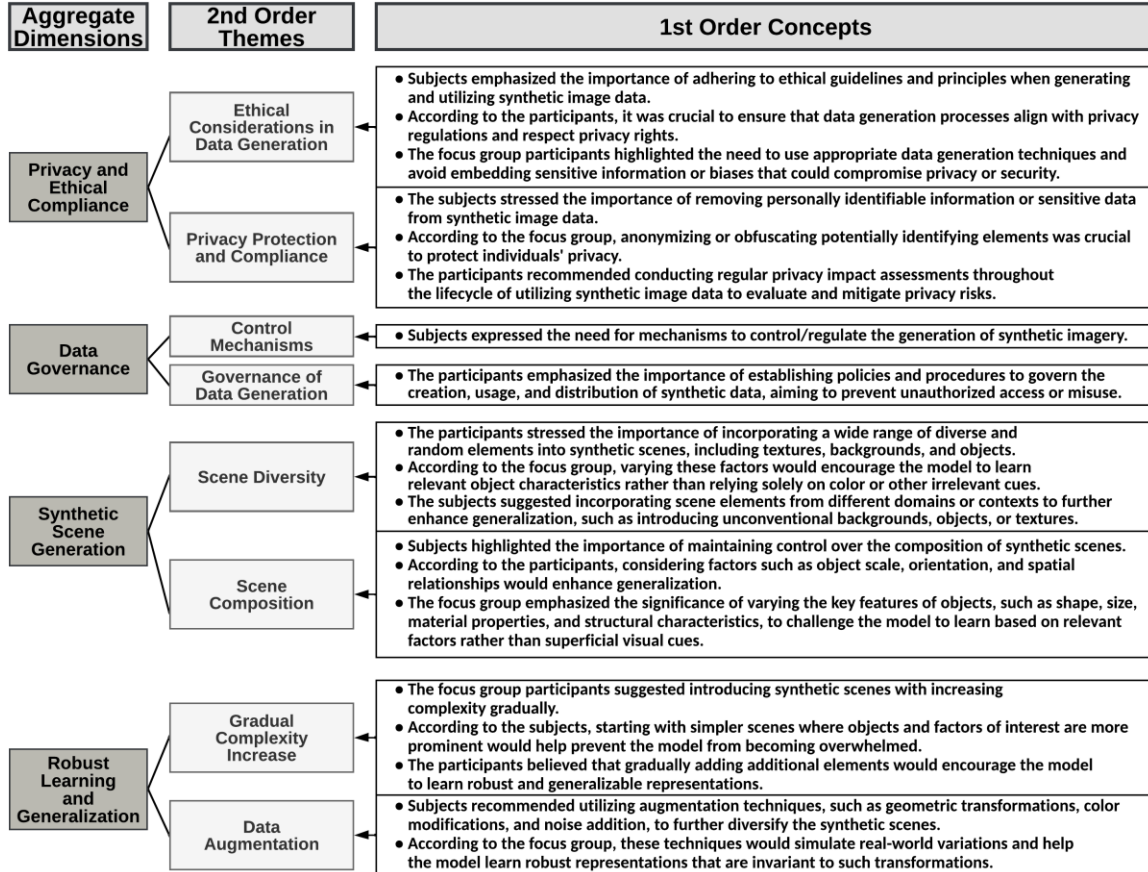


Figure 2: Focus group results

The conducted focus group consisted of n=11 participants, including three AI senior scholars, two AI research associates, two computer vision project leads, and four IS researchers, all with professional experience ranging from 3-17 years. The goal of the moderated focus group was to develop design knowledge (e.g., user-specific requirements, characteristics, process steps), but without incorporating the literature findings to avoid any bias. To ensure qualitative rigor during and after this session, the well-established methodology outlined by Gioia et al. [35] was followed throughout, which involved formulating *first order concepts*, *second order themes*, and *aggregate dimensions (AD)* based on the subjects' expressed statements. The focus group findings are shown in Figure 2. The focus group findings revealed key insights regarding the utilization of synthetic image data in computer vision deep learning settings. Participants emphasized the importance of adhering to ethical guidelines and privacy regulations throughout the data generation process. They also stressed the need to remove personally identifiable information and conduct privacy impact assessments regularly to mitigate privacy risks - resulting in **AD1** (Privacy and Ethical Compliance). The subjects also highlighted the significance of implementing mechanisms for generation control and data governance to prevent unauthorized access or misuse which was epitomized by **AD2** (Data Governance). Additionally, the focus group emphasized the need for synthetic scene diversity, recommending the incorporation of various elements and cross-

domain scene randomization to enhance generalization. While promoting scene diversity, they emphasized the importance of maintaining control over scene composition to ensure proper representation of intended features and factors of interest. This resulted in **AD3** (Synthetic Scene Generation). The participants further suggested gradually increasing the complexity of synthetic scenes to prevent overfitting and promote robust learning and generalization. They also recommended data augmentation techniques, such as geometric transformations and color modifications, to diversify synthetic scenes and enable the learning of robust representations invariant to real-world variations – illustrated by **AD4** (Robust Learning and Generalization).

Overall, these focus group findings provided valuable insights for the proposed design principles, which aim to guide the ethical and effective utilization of synthetic image data in computer vision research and applications.

4. Design Principles for Using Synthetic Image Data

As aforementioned, the performed design cycle was dedicated to creating design knowledge and developing theoretically sound design principles as the main artifact. As design principles embody a general design solution for a class of problems [17], they are of prescriptive and universal nature, specifying how a solution should be designed to achieve the desired objective [36]. In this context, the DPs were derived from a supportive approach and the conceptual schema of Gregor et al. [18], whose a priori specification suggests prescriptive wording [36], thereby allowing us to formulate accessible, precise, and expressive design knowledge, as elucidated by the framework [18]. In addition to utilizing the *anatomy of a design principle* [18] and the kernel theory of *value sensitive design* [12], the development and wording of the design principles were guided by the ethical design science research framework proposed by Durani et al. [16], resulting in more prescriptive guidance for leveraging the positive impact of the artifact and minimizing its adverse effects.

The design principles were rigorously derived from the literature and the aggregate dimensions from the qualitatively analyzed focus group results. As shown in Figure 3, seven specific design principles were developed to address the identified problem of lacking design knowledge on how to utilize synthetic image data in computer vision deep learning environments.

DP1 draws from *AD1* and states that ethical guidelines and principles should be followed when generating and utilizing synthetic image data. Incorporating value sensitive design [12], it is important to align data generation processes with privacy regulations and to show respect for individual privacy rights. Therefore, care should be taken to employ suitable data generation techniques (e.g. via *Unity3D*) and to refrain from incorporating sensitive information or biases that could potentially compromise the privacy or security of individuals.

DP2 also builds on *AD1* and addresses the need for the synthetic image data to contain no personally identifiable information (PII) or sensitive data. It's necessary to anonymize or obfuscate any elements that could potentially reveal an individual's identity. Throughout the process of using synthetic imagery, regular privacy impact assessments should be conducted to assess the privacy risks associated with the generation, storage, transmission, and use of the data. Appropriate measures should then be implemented to mitigate the identified risks and ensure ongoing compliance with privacy regulations, thus aligning with value sensitive design [12]. In this regard, it is recommended that differential privacy mechanisms be incorporated into the generation and use of synthetic image data, where controlled noise or perturbations are introduced during data generation to prevent individual data points from being distinguished with a high degree of certainty [3, 4]. This approach can protect the privacy of individuals even in the presence of external information.

Based on *AD2*, **DP3** states that mechanisms should be implemented to control and regulate the generation of synthetic image data, such as process frameworks, toolkits, virtual environments, or guidelines. Hence, policies and procedures need to be established to govern the creation, usage,

and distribution of synthetic data in order to prevent unauthorized access or misuse, ensuring value sensitive design [12].

DP4 stems from *AD3* and specifies that a wide range of diverse and random elements should be incorporated into synthetic scenes, including textures, backgrounds, and objects. By varying these factors, the model will be encouraged to learn relevant object characteristics instead of relying on color or other irrelevant cues [33, 3]. To further improve generalization, cross-domain scene randomization should be used, which involves incorporating scene elements from different domains or contexts (e.g., non-healthcare elements in healthcare settings). Introducing unconventional backgrounds, objects, or textures that are not typically associated with the objects of interest can push the model to learn their intrinsic properties, thereby promoting adaptability to real-world scenarios [3, 28].

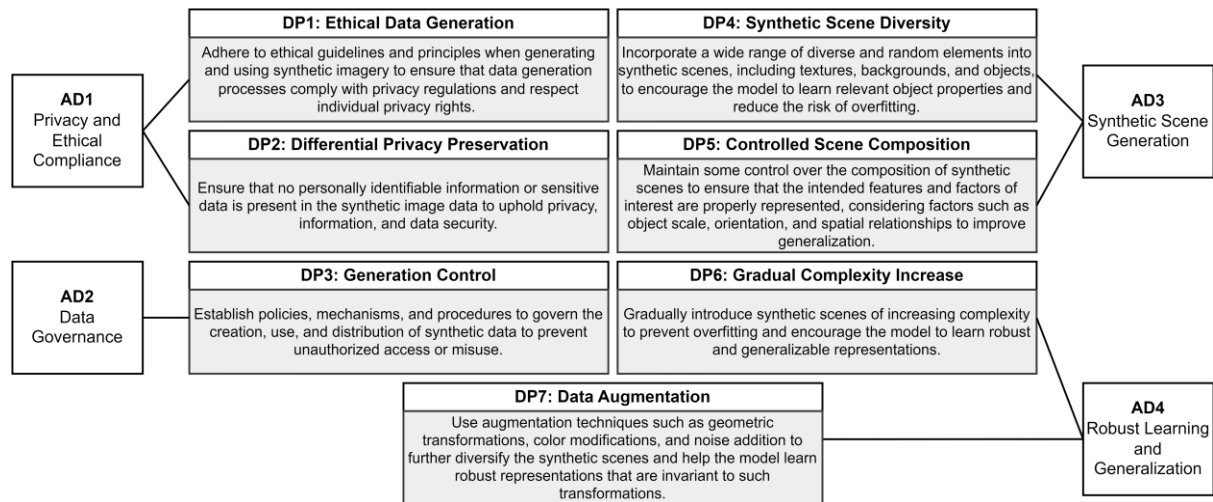


Figure 3: Design principles for using synthetic image data in computer vision

Note: DP = design principle; AD = aggregate dimension

DP5 closely connects to **DP4** and further relates to *AD3*, stating that while aiming to promote scene diversity (and randomness), it is important to maintain a level of control over the composition of synthetic scenes. This ensures that the intended features and factors of interest are properly represented where factors such as object scale, orientation, and spatial relationships should be considered to enhance generalization [6, 33]. Rather than relying solely on changing the appearance of objects, the focus should be on varying their key features, and changing attributes such as shape, size, material properties, and structural characteristics will challenge the model to learn object representations based on these relevant factors rather than superficial visual cues.

DP6 draws from *AD4* and addresses the gradual introduction of synthetic scenes with increasing complexity. Training the deep learning model should begin with simpler scenes that highlight objects and factors of interest more prominently, and gradually incorporate additional elements to prevent overwhelming the model [1]. This approach helps prevent overfitting and encourages the model to learn robust and generalizable representations, which is highly relevant when working with synthetic rather than real data [5, 3, 29].

DP7 also stems from *AD4* and states that augmentation techniques, such as geometric transformations, color modifications, and noise addition, should be utilized to enhance the diversity of synthetic scenes [31, 3, 4]. These techniques simulate real-world variations and assist the model in learning robust representations that remain invariant to such transformations, which further mitigates the risk of model overfitting [1].

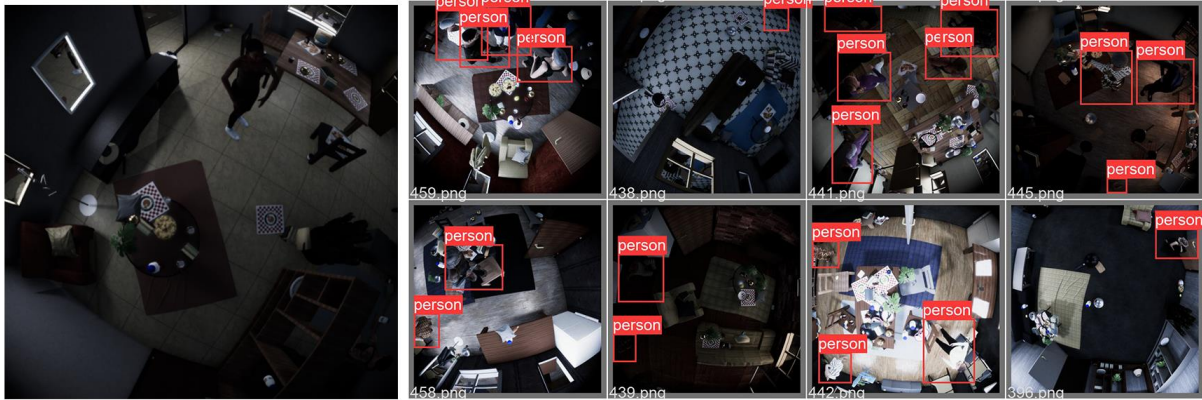


Figure 4: Exemplary synthetic image data instantiation

Figure 4 shows an exemplary instantiation of the design principles in a computer vision deep learning setting for person detection, which is therefore concerned with ethical data generation. By following value sensitive design [12], abstract and non-genuine characters were generated for the scenes, ensuring ethical data generation (DP1) and differential privacy preservation (DP2). Using a video game engine and various mechanisms to generate the synthetic image data (DP3), a wide variety of scenes (DP4) and compositions (DP5) were achieved, significantly and intentionally varying the key features. In addition, several scenes are epitomized by reduced complexity (DP6) to achieve better generalization, where this can be scaled infinitely. Finally, data augmentation techniques (DP7) in the form of geometric transformations (i.e., lens distortion) were used to reduce the risk of overfitting. By incorporating these design principles, the depicted figure ensures ethical data generation practices in person detection, promoting accurate and reliable results while considering privacy, diversity, generalization, and controlled complexity.

5. Evaluation

To ensure scientific rigor in the evaluation of this design cycle, the well-established *FEDS* framework proposed by Venable et al. [37] was used. the evaluation phase is highly relevant in design science research [11, 37], it is necessary to select an appropriate strategic process and determine the constructs to be evaluated. Given the small and rather simple design of the main artifact, embodied in a set of design principles that result in low social and technical risk and uncertainty, the evaluation strategy of *quick & simple* [37] was chosen. Thus, the goal was to conduct an evaluation episode to complete the design cycle and to move quickly to a summative evaluation. The evaluation schema employed in this study takes into account the roles of key stakeholders involved the formulation of design principles. This schema allows design science researchers to assess the usability of generated design principles for different user groups. Two critical questions arise from this perspective: first, whether the design principles are understandable and useful to implementers, and second, whether they effectively serve the goals of users who implement the resulting instantiations [18]. Therefore, *evaluation activity 2* [38] was used, which describes an artificial activity, since the artifact has not yet been properly instantiated (note that Figure 4 serves only as an exemplary visualization). This activity aims to validate the principles of form and function, which have been developed during the design cycle [38].

Table 1
Think aloud results and reusability framework application

Reusability Category	Verbalized Think Aloud Results
<i>Accessibility</i>	<p>The participants stated that they found the design principles to be highly accessible, emphasizing the clarity and understandability of the language used. Particularly in terms of privacy and data protection (DP1, DP2, DP3), the participants recognized the importance of practitioners being able to comprehend and implement these principles effectively, ensuring ethical and privacy-compliant use of synthetic image data.</p> <p>However, participants noted that while the design principles presented were well constructed, there was a suggestion to consider including explanations for technical terms such as "differential privacy mechanisms". They mentioned that providing brief definitions for such terms could help readers who may not be deeply familiar with the field to better understand the content.</p>
<i>Importance</i>	<p>The participants also highlighted the significant importance of the design principles. They acknowledged that these principles addressed crucial concerns related to privacy, data anonymization, information security, and regulatory compliance (DP1, DP2, DP3). By incorporating these principles into deep learning environments, the participants emphasized the practical relevance and significance of adhering to them, fostering trust and responsible use of synthetic image data.</p>
<i>Novelty and Insightfulness</i>	<p>The participants expressed their appreciation for the design principles, stating that they introduced fresh perspectives to the generation and utilization of synthetic image data.</p> <p>The emphasis on diversity and randomness in scene composition, the incorporation of unconventional elements, and cross-domain randomization were noted as innovative approaches (DP4, DP5). According to the participants, these principles challenged traditional methods and encouraged thinking beyond the conventional, promoting adaptability to real-world scenarios.</p>
<i>Actability and Guidance</i>	<p>The participants commended the actability and appropriate guidance provided by the design principles. They highlighted the clear frameworks, policies, and mechanisms suggested to regulate the generation and usage of synthetic image data (DP3). The participants found the gradual introduction of complexity during training and the utilization of augmentation techniques as practical suggestions aligned with their deep learning workflows (DP6, DP7). The guidance provided struck a balance between providing direction and allowing for creative application of the principles.</p>
<i>Effectiveness</i>	<p>The effectiveness of the design principles was evident to the participants. They recognized the emphasis on preventing privacy risks (DP1, DP2), mitigating overfitting, and enhancing model generalization (DP4, DP5, DP6, DP7). By adhering to these principles, the participants noted that robust deep learning models could be developed, yielding high performance on real-world data.</p> <p>They found the strategies of gradually introducing complexity (DP6) and using augmentation techniques (DP7) to be effective in optimizing performance and ensuring the practical utility of synthetic image data.</p> <p>However, participants expressed that while the concepts of DP6 and DP7 were intriguing, they suggested that a comparative analysis be included which could contrast the proposed principles with existing methodologies and highlight the unique advantages and improvements.</p>

To ensure the objectives of feasibility, accessibility, completeness, and applicability, it seems reasonable to apply the framework of design principle reusability proposed by Iivari et al. [39]. This framework provides a systematic approach to evaluating the design principles generated during the design cycle and, by assessing the reusability of these principles, researchers can determine their potential for wider application and adoption in similar contexts [39]. For this purpose, and in accordance to the kernel theory of value sensitive design [12], a qualitative think aloud session addressing the reusability of the proposed design principles was conducted. Therefore, the method of *concurrent think-aloud* [40] was employed with n=9 AI experts, where the sample size was decided based on the “10±2 rule” for think aloud sessions [41]. The participants were asked to verbalize their thoughts about the design principles in terms of the reusability categories proposed by Iivari et al. [39]. The experts were provided with a detailed textual description of the design principles, along with visual examples of synthetic image data. Table 1 presents the qualitative think aloud results, including the categories of the reusability framework and the clustered verbalized thoughts of the participants.

Overall, the participants of the think aloud session positively evaluated the design principles, emphasizing their accessibility, importance, novelty and insightfulness, actability with appropriate guidance, and overall effectiveness in enhancing the use of synthetic image data in deep learning environments. Their feedback underscored the value and especially the reusability of these principles in guiding practices and ensuring responsible and efficient utilization of synthetic image data in deep learning. Nonetheless, a few areas for improvement have been identified by the participants. However, a number of potential areas for refinement emerged from their constructive feedback. Participants noted that while the structure of the design principles was commendable, they suggested that clarifications of technical terms such as “Differential Privacy Mechanisms” could be included. It was suggested that providing concise definitions for these terms could serve to help readers or researchers less familiar with the field to better understand the content. In addition, participants expressed the notion that despite the appeal of *DP6 (Gradual Complexity Increase)* and *DP7 (Data Augmentation)*, it might be prudent to introduce a comparative analysis that could compare the proposed principles with existing methodologies, thereby highlighting their particular merits and improvements. These suggestions for refinement, which come from the participants and should be picked up in subsequent design science cycles, are intended to increase the accessibility and effectiveness of the design principles, serve a wider range of readers, and further substantiate their utility.

6. Conclusion

The paper proposes general design principles for the use of synthetic image data in computer vision deep learning environments to ensure more ethical, robust, traceable, and effective development and implementation of such models. Consequently, to answer the initially formulated research question of this paper, the results of a completed design science research cycle have been presented. Hereby, the positive evaluation of the design principles substantiates the theoretical and practical relevance of the design principles and researchers can adapt these to develop, utilize, or modify deep learning models based on synthetic image data. By using the DSR paradigm [10], the study moves beyond descriptive knowledge and aims to provide prescriptive knowledge, focusing on the design principles for utilizing synthetic image data in deep learning. This integration of the design science research paradigm contributes to the advancement of design knowledge in the field along with the IS design science knowledge base according to Woo et al. [42]. The paper also contributes theoretically by employing the value sensitive design theory, as proposed by Friedman et al. [12], which enhances the understanding of the ethical implications and user values in the context of synthetic image data utilization. The practical implications of these design principles include improved performance, enhanced privacy protection, and responsible and efficient utilization of synthetic image data in real-world applications, while the reusability of these principles in similar contexts contributes to their

wider application and adoption, promoting responsible and efficient utilization of synthetic image data in computer vision.

Meanwhile, in the context of the positive evaluation episode, the following limitations should be considered: First, design principles and their development are tied to the subjective creativity of the researcher, even after various data collection episodes and literature reviews. However, not all design decisions can or should be derived from behavioral or mathematical theories, as some degree of creativity is essential to developing an innovative design artifact [43, 44], whereas a certain degree of rigor can be implemented such as the utilized methodological approaches of Gregor et al. [18], Möller et al. [19], or Fu et al. [36]. Second, as with any other evaluation, the results describe only one sample, meaning that different results could be expected if a different sample were chosen. Therefore, this particular limitation could be addressed in future research, while the application of the design principles (in various domains such as digital health, etc.) as part of a case study or framing guideline seems highly interesting. It would be presumptuous to assume that the design principles contain all the necessary information that will need to be either refined, adapted, or expanded in future research efforts. Moreover, the highlighted areas for improvement of the design principles based on the reusability framework could be addressed in a subsequent design science cycle and future research.

References

- [1] L. Alzubaidi, J. Zhang, et al., Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions, *Journal of Big Data* 8 (2021), 1-74.
- [2] S. Hinterstoisser, O. Pauly et al., An annotation saved is an annotation earned: Using fully synthetic training for object detection in 'Proceedings of the IEEE/CVF' (2019).
- [3] V. Seib, B. Lange, S. Wirtz, Mixing Real and Synthetic Data to Enhance Neural Network Training - A Review of Current Approaches, 2020, arXiv:2007.08781.
- [4] X. Zhang, Y. Fu, S. Jiang, X. Xue, Y.G. Jiang, G. Agam, Stacked multichannel autoencoder—an efficient way of learning from synthetic data. *Multimedia Tools and Applications* 77 (2018), 26563-26580.
- [5] J.J. Bird, D.R. Faria, A. Ekárt, P.P. Ayrosa, From simulation to reality: CNN transfer learning for scene classification, in 2020 IEEE 10th International Conference on Intelligent Systems (2020), IEEE, 619-625.
- [6] M. Krump, M. Ruß, P. Stütz, Deep learning algorithms for vehicle detection on UAV platforms: first investigations on the effects of synthetic training, in *Modelling and Simulation for Autonomous Systems: 6th International Conference, MESAS 2020, Palermo, Italy*, 50-70.
- [7] F. Kuhnke, J. Ostermann, Deep head pose estimation using synthetic images and partial adversarial domain adaption for continuous label spaces, in *Proceedings of the IEEE/CVF International Conference on computer vision (2019)*, 10164-10173.
- [8] J. Liu, H. Rahmani, N. Akhtar, A. Mian, Learning human pose models from synthesized data for robust RGB-D action recognition. *International Journal of Computer Vision* 127 (2019), 1545-1564.
- [9] C. Taleb, L. Likforman-Sulem, C. Mokbel, Improving deep learning Parkinson's disease detection through data augmentation training, in *Pattern Recognition and Artificial Intelligence: Third Mediterranean Conference, MedPRAI, 2020, Istanbul, Turkey*, 79-93.
- [10] A. Hevner, S. March, J. Park, S. Ram, Design science in information systems re-search, *MIS Quarterly* 28(1) (2004), 75-105.
- [11] V.K. Vaishnavi, W. Kuechler, *Design science research methods and patterns: innovating information and communication technology*, 2015, Crc Press.
- [12] B. Friedman, P.H. Kahn, A. Borning, A. Hultdtgren, *Value sensitive design and information systems. Early engagement and new technologies: Opening up the laboratory*, 2013, 55-95.
- [13] S. Gregor, A.R. Hevner, Positioning and presenting design science research for maximum impact. *MIS Quarterly* (2013), 337-355.

- [14] J. Mingers, J. Brocklesby, Multimethodology: Towards a framework for mixing methodologies. *Omega*, 25(5), 1997, 489-509.
- [15] J.R. Venable, J. Pries-Heje, R.L. Baskerville, Choosing a design science research methodology in ACIS 2017 Proceedings, 2017, 112.
- [16] K. Durani, A. Eckhardt, T. Kollmer, Towards ethical design science research in ICIS 2021 Proceedings, 2021, 3.
- [17] R. Baskerville, J. Pries-Heje, Explanatory design theory. *Business & Information Systems Engineering*, 2 (2010), 271-282.
- [18] S. Gregor, L. Chandra Kruse, S. Seidel, Research perspectives: the anatomy of a design principle. *Journal of the Association for Information Systems*, 2020.
- [19] F. Möller, T. Schoormann, G. Strobel, M.R.P. Hansen, Unveiling the Cloak: Kernel Theory Use in Design Science Research in ICIS 2022 Proceedings, 2022.
- [20] F. Niederman, S. March, The “theoretical lens” concept: We all know what it means, but do we all know the same thing? *Communications of the Association for Information Systems*, 44(1), 2019, 1.
- [21] S. Lee, B. Park, A. Kim, Deep learning based object detection via style-transferred underwater sonar images. *IFAC-PapersOnLine*, 52(21), 2019, 152-155.
- [22] T. Body, X. Tao, Y. Li, L. Li,, N. Zhong, Using back-and-forth translation to create artificial augmented textual data for sentiment analysis models. *Expert Systems with Applications*, 178, 2021, 115033.
- [23] F. Condrea, V.A. Ivan, M. Leordeanu, In search of life: Learning from synthetic data to detect vital signs in videos, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, 298-299.
- [24] H.F. Zaki, F. Shafait, A. Mian, Viewpoint invariant semantic object and scene categorization with RGB-D sensors. *Autonomous Robots*, 43 (2019), 1005-1022.
- [25] A. Lahiri, A. Agarwalla, P.K. Biswas, Unsupervised domain adaptation for learning eye gaze from a million synthetic images: An adversarial approach, in *Proceedings of the 11th Indian Conference on Computer Vision (2018), Graphics and Image Processing*, 1-9.
- [26] H. Venkateswara, S. Chakraborty, S. Panchanathan, Deep-learning systems for domain adaptation in computer vision: Learning transferable feature representations. *IEEE Signal Processing Magazine*, 34(6), 2017, 117-129.
- [27] N. Aranjuelo, S. García, E. Loyo, L. Unzueta, O. Otaegui, Key strategies for synthetic data generation for training intelligent systems based on people detection from omnidirectional cameras. *Computers & Electrical Engineering*, 92 (2021), 107105.
- [28] S.Z. Valtchev, J. Wu, Domain randomization for neural network classification. *Journal of big Data*, 8(1), 2021, 94.
- [29] L. Wan, J. Zhou, B. Zhang, Data Synthesis for Document Layout Analysis, in *International Symposium on Emerging Technologies for Education*, 2020, 244-252.
- [30] H. Abu Alhaija, S.K. Mustikovela, L. Mescheder, A. Geiger, C. Rother, Augmented reality meets computer vision: Efficient data generation for urban driving scenes. *International Journal of Computer Vision*, 126, 2018, 961-972.
- [31] M. Müller, V. Casser, J. Lahoud, N. Smith, B. Ghanem, Sim4cv: A photo-realistic simulator for computer vision applications. *International Journal of Computer Vision*, 126, 2018, 902-919.
- [32] M. Valerio Giuffrida, H. Scharr, S.A. Tsiftaris, Arigan: Synthetic arabidopsis plants using generative adversarial network, in *Proceedings of the IEEE international conference on computer vision workshops*, 2017, 2064-2071.
- [33] T. Scheck, R. Seidel, G. Hirtz, Learning from theodore: A synthetic omnidirectional top-view indoor dataset for deep transfer learning, in *Proceedings of the IEEE/CVF Winter conference on applications of computer vision (2020)*, 943-952.
- [34] D.L. Morgan, *Qualitative Research Methods: Focus groups as qualitative research (2)*, 1997, Thousand Oaks SAGE Publications, Inc.
- [35] D.A. Gioia, K.G. Corley, A.L. Hamilton, Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*, 16(1), 2012, 15-31.

- [36] K.K. Fu, M.C. Yang, K.L. Wood, Design principles: Literature review, analysis, and future directions. *Journal of Mechanical Design*, 138(10), 2016, 101103.
- [37] J. Venable, J. Pries-Heje, R. Baskerville, FEDS: a framework for evaluation in design science research. *European Journal of Information Systems*, 25 (2016), 77-89.
- [38] C. Sonnenberg, J. Vom Brocke, Evaluation patterns for design science research artefacts, in *Practical Aspects of Design Science: European Design Science Symposium (2012)*, Leixlip, Ireland, 71-83.
- [39] J. Iivari, M.R.P. Hansen, A. Haj-Bolouri, A framework for light reusability evaluation of design principles in design science research, in *13th International Conference on Design Science Research and Information Systems and Technology: Designing for a Digital and Globalized World*, 2018.
- [40] M. Van Den Haak, M. De Jong, P. Jan Schellens, Retrospective vs. concurrent think-aloud protocols: testing the usability of an online library catalogue. *Behaviour & Information Technology* (2003), 22(5):339-351.
- [41] W. Hwang, G. Salvendy, Number of people required for usability evaluation: the 10±2 rule. *Communications of the ACM* (2010) 53(5):130-133.
- [42] C. Woo, A. Saghafi, A. Rosales, What is a Contribution to IS Design Science Knowledge?, in *Thirty Fifth International Conference on Information Systems*, 2014, Auckland.
- [43] A. Hevner, S. Chatterjee, *Design science research in information systems*, Design research in information systems, 2010, Springer, Boston, 9-22.
- [44] R. Baskerville, M. Kaul, J. Pries-Heje, V.C. Storey, E. Kristiansen, Bounded creativity in design science research, in *ICIS 2016 Proceedings*.

Privacy-Enhancing Technologies in the Process of Data Privacy Compliance: An Educational Perspective

Alexandra Klymenko, Stephen Meisenbacher, Florian Messmer and Florian Matthes

Technical University of Munich, School of Computation, Information and Technology, Department of Computer Science, Boltzmannstr. 3, Garching, 85748, Germany

Abstract

Achieving data privacy compliance presents a unique interdisciplinary challenge for experts from many backgrounds, particularly the technical and legal professions. As a potential solution for the legal mandate handed down by modern privacy regulations, Privacy-Enhancing Technologies (PETs) can serve as promising tools to help data processors demonstrate compliance. The implementation of PETs does not come immediately, however, and challenges in their adoption include their inherent technical complexity, as well as the lack of awareness and understanding of these technologies. In tackling these challenges, we investigate the educational needs of practitioners working in privacy compliance. Guided by Bloom's Revised Taxonomy, we begin the discussion on how the adoption of PETs can become more informed, with the goal of improving the efficiency and privacy consciousness of compliance programs. To accomplish this, we conduct 11 semi-structured interviews, analyze the results following Grounded Theory, and evaluate our findings in a survey with 24 respondents.

Keywords

Data privacy, privacy compliance, privacy-enhancing technologies, continuous education

1. Introduction

In a world where vast amounts of data are being created and processed on a continual basis, the need for the responsible handling of such data has starkly risen. Along with increasing concerns regarding the protection of individuals' privacy, the pressure placed on practitioners to comply with relevant data privacy regulations such as the GDPR raises the stakes for data processors [1][2]. Ultimately, a technical response in the form of privacy preservation must be implemented in data-intensive systems, a complex task that is accompanied by multiple challenges [3].

Recently, the promise of Privacy-Enhancing Technologies (PETs) has saturated the academic sphere, engaging researchers to develop innovative technologies for data privacy protection. In essence, PETs encompass a range of technical approaches designed to protect the data of the individual, when this data is utilized for some purpose. Such technologies, while falling under the same class of *Privacy-Enhancing Technologies*, are highly diverse, particularly in their applicable use cases. One unifying aspect, though, is their inherent complexity, which has kept their practical adoption quite limited [4][5]. Nevertheless, data processors can benefit from the deployment of PETs as a means of protecting sensitive information while still allowing meaningful utilization of the data.

The road to widespread adoption of Privacy-Enhancing Technologies begins with the transition from PETs as a research topic to the dissemination of such knowledge to practitioners in the industry. However, essential questions then arise as to who constitutes the target audience, and what specific knowledge regarding PETs is required by practitioners. To identify the target

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ alexandra.klymenko@tum.de (A. Klymenko); stephen.meisenbacher@tum.de (S. Meisenbacher); f.messmer@tum.de (F. Messmer); matthes@tum.de (F. Matthes)

ORCID 0000-0001-7485-2933 (A. Klymenko); 0000-0001-9230-5001 (S. Meisenbacher); 0000-0002-6667-5452 (F. Matthes)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

audience, we look to the process of privacy compliance, which centers around the implementation of appropriate technical measures for the safeguarding of personal data being processed in a system. Gürses and Del Alamo [6] and Klymenko et al. [5] have shown that this process is highly interdisciplinary, involving primarily experts of technical and legal backgrounds. These two types of roles, therefore, become the focus of our work. We argue that education on PETs should take into consideration the diversity of roles in the privacy compliance process, as differing roles have distinct backgrounds, responsibilities, concerns, and, as will be shown, different interests regarding familiarization with PETs.

In this work, we aim to investigate the educational needs of practitioners with respect to PETs, with the goal of empowering them to be competent users of PETs, as "computer scientists and particularly IT security experts with knowledge about privacy-enhancing technologies are increasingly needed" [7]. We define the following research questions:

[RQ1] How can learning goals for Privacy-Enhancing Technologies be defined?

[RQ2] How can these learning goals be mapped to the role-specific needs of practitioners involved in privacy compliance?

To answer these research questions, we draw upon existing educational frameworks, leveraging the resulting insights from industry interviews to augment educational thinking on PETs. The possible learning objectives with regards to Privacy-Enhancing Technologies are segmented according to the framework of Bloom's Revised Taxonomy [8] introduced by Krathwohl. Subsequently, we evaluate the identified objectives via the administration of surveys. From this, we propose a new way of thinking about education on PETs, particularly considering the background of the person in question.

2. Background

A key step towards ensuring compliance with the data privacy regulations comes with the requirement to implement technical measures to protect the privacy of individuals. In this respect becomes important the concept of Privacy-Enhancing Technologies (PETs), a class of technologies that "protect privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system" [9]. The recent guidance by the Information Commissioner's Office (ICO) provides a detailed discussion on some of the prominent PETs, such as Differential Privacy, Zero-Knowledge Proofs, and Secure Multi-Party Computation, and outlines how they can help organizations to achieve data privacy compliance [10].

Although such advanced PETs present concrete solutions for personal data protection and multiple real-world use case examples have been reported [11][12], they still remain predominantly in the academic sphere and are not widely adopted in practice [4][5]. Among the main reasons for this, is the complexity of these technologies, as well as the lack of awareness, knowledge, and education on them [3]. Therefore, the promotion of continuing education on topics related to data privacy and PETs can be considered crucial to the development of successful privacy compliance programs. While the presented recent reports [10][11][12] highlight the significance of PETs and play an important role in promoting their implementation in the industry, these works offer a rather broader overview and do not focus on providing tailored and comprehensive educational content.

In this work, we consider the inherently interdisciplinary nature of privacy compliance and investigate the educational needs of practitioners based on the different roles involved in the process of privacy compliance, as proposed by Klymenko et al. [5]. Namely, the roles are divided into three categories: 1) *Legal* – practicing lawyers, specializing in the fields of privacy and data protection, 2) *Technical* – roles involved in the implementation of the product, such as software developers, engineers, and architects, as well as the appropriate management roles, and 3) *Go-Betweens* – practitioners working at the intersection of technical and legal fields, including roles such as Data Protection Officer (DPO), and Privacy Engineer.

3. Methodology

To assess the educational needs of practitioners in learning about PETs, we designed the interview and survey studies to focus on extracting the *learning goals* of the questioned experts. In the interview, this was done in a semi-structured way, with two categories of questions: background questions, including the interviewee’s baseline knowledge of PETs, and questions aimed at identifying what kind of information about PETs is most relevant to the interviewee’s role and responsibilities. A thematic content analysis according to Braun and Clarke [13] was conducted on the interview transcripts. The main goal of this analysis was to identify overarching themes expressed in the interviews, particularly relating to the learning needs and goals of practitioners with respect to PETs. Guided by following Grounded Theory (GT) Methodology [14], we analyzed interview transcripts concurrently to data collection and highlighted key themes, which were categorized into learning goals and educational needs. Axial coding was applied to identify relationships between these themes, supported by Bloom’s Revised Taxonomy.

Based on the resulting learning goals identified by our analysis and introduced in Table 3, the survey statements were designed to map learning goals to role-specific educational needs, where each statement corresponded to a cognitive process in Bloom’s Revised Taxonomy. The survey participants were then prompted to select the statement which best reflects their personal learning goals, allowing for the mapping of roles to levels in Bloom’s Revised Taxonomy.

Table 1 and Table 2 present relevant information on the interviewees and survey participants. *Area* identifies whether the survey respondents are working in a technical (T), legal (L), or Go-Between (G) role. *Exp.* represents years of relevant experience. To mitigate bias, no survey respondents also took part in the interview study.

Table 1
Interview study participants

ID	Role	Area	Exp.	Sector
IP01	Product Owner	T	5	Machinery
IP02	CSO / Co-Founder	T	2	Software Development
IP03	System Administrator	T	2	Electronics Manufacturing
IP04	Trainee IT Strategy	T	1	Automotive
IP05	IAM Architect	T	6	Electronics Manufacturing
IP06	Solution Architect	T	5	Machinery
IP07	CTO, Co-Founder	T	4	Software Development
IP08	GDPR Senior Data Privacy Ambassador	G	33	Health Services
IP09	Developer, Owner	T	8	Software Development
IP010	Head, Applied Privacy Technologies Group	T	10	IT Services
IP11	Researcher, Applied Privacy Technologies Group	T	5	IT Services

4. PETs and Bloom's Revised Taxonomy

To formulate and categorize the learning goals of practitioners regarding PETs, we employ Bloom's Revised Taxonomy [8]. This taxonomy provides an organizational structure of educational objectives, consisting of the *Knowledge Dimension* and the *Cognitive Process Dimension*. The types of knowledge are structured into four categories: *Factual*, *Conceptual*, *Procedural*, and *Metacognitive*. These knowledge levels are mapped back to the six cognitive processes: *Remember*, *Understand*, *Apply*, *Analyze*, *Evaluate*, and *Create*. The knowledge levels are introduced below in light of PETs, and the resulting learning goal statements are presented in Table 3, which maps statements to their corresponding knowledge level and cognitive process. This mapping becomes relevant to understanding the role-specific learning goals of practitioners on the topic of PETs.

Table 2
Survey study participants

ID	Role	Area	Exp.	Sector
SP01	External Consultant (Law)	L	3-5	Health services
SP02	Developer	T	<3	Health services
SP03	Developer	T	3-5	Engineering
SP04	Developer	T	<3	Financial services
SP05	Developer	T	<3	Engineering
SP06	Legal Counsel	L	5-10	Financial services
SP07	Project Owner	T	<3	Media
SP08	Architect	T	3-5	Engineering
SP09	Data Protection Officer	G	5-10	Construction
SP010	Developer	T	<3	N/A
SP11	Developer	T	3-5	Engineering
SP12	Management	T	3-5	Financial service
SP13	Privacy Engineer	G	3-5	Public service
SP14	Compliance Officer	L	<3	N/A
SP15	Legal Counsel	L	<3	Financial services
SP16	External Consultant (Law)	L	5-10	N/A
SP17	Project Manager	T	<3	Financial services
SP18	Developer	T	<3	Financial services
SP19	Project Manager	T	10-20	Engineering
SP20	Architect	T	5-10	Education
SP21	Privacy Engineer	G	<3	Education
SP22	Legal Counsel	L	5-10	Public service
SP23	IT Architect	T	10-20	Public service
SP24	Management	T	20+	Media

4.1. Factual Knowledge

Factual knowledge includes terminology, characteristics, and features of PETs. The simplest learning goals are to list different PETs, as well as to know about the use cases of PETs, a topic most directly corresponding to *Remember*. Analyzing PETs on a factual level can be conceived as comparing different PETs and accordingly selecting technologies. It thus becomes clear that the tasks build up on each other, i.e., that *Remember*, *Understand*, and *Apply* are required to perform the subsequent *Analyze* tasks.

4.2. Conceptual Knowledge

Conceptual knowledge is closely related to theoretical topics, such as introducing models, approaches, and interrelations of PETs. As opposed to factual knowledge, conceptual knowledge includes the principles behind the functionality of PETs. Based on the study results, statements are focused on system architecture, as interview participants reported a need to understand this topic better. The idea of integrating newly learned information into existing knowledge domains characterizes conceptual knowledge. However, it encapsulates the decision over which technology would be applicable; the implementation itself belongs strictly to the following category.

4.3. Procedural Knowledge

Here, the focus is placed on the implementation of PETs. Although the statements presented are expected to be universally applicable to all privacy roles, there is now a shift towards more technical content. Applying Procedural Knowledge marks the point where the learning content becomes rather technical, implying that a higher level of technical literacy is required. Furthermore, it shows how many learning goals can be identified before implementation. The next modification of the cognitive category is directed at the implementation action itself. The intent is not just to implement PETs in any fashion but to know parameters and quality measures, and thereby build an implementation strategy. Ultimately, the goal of procedural knowledge is not only to find the most suitable PET, but also to contribute to the development of new PETs.

4.4. Metacognitive Knowledge

Privacy-Enhancing Technologies are under constant pressure to evolve, as are any technologies employed to minimize risks or mitigate threats. The question of maturity is of great interest with Privacy-Enhancing Technologies. Achieving such knowledge requires a deep knowledge of the PETs in question, the environment in which PETs are implemented, and awareness of the limitations of the technologies. Therefore, learning goals in this knowledge category convey this critical approach, while also focusing on finding strategies to address these limitations. The highest learning goal would be to transfer knowledge to formerly unknown domains, identifying new purposes for PETs.

4.5. Learning Goal Statements

Table 3 presents the set of learning goal statements for PETs, which is based on Bloom’s Revised Taxonomy and supported by the interview findings. Using the guidelines provided by the original taxonomy and augmenting these with goals expressed by interviewees, we build the statements in Table 3 to align with the knowledge levels and cognitive processes of Bloom’s Revised Taxonomy. This mapping process is aided by Anderson and Krathwohl [15] and inspired by Servin et al. [16], the latter of which extends existing verb sets to include the technical domain.

Table 3
Learning goal statements based on Bloom’s Revised Taxonomy

	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge	I want to know what different PETs exist.	I want to know the various use cases for PETs.	I want to be able to follow discussions about PETs.	I want to be able to differentiate PETs.	I want to verify statements about the features of PETs.	I want to be able to classify a new PET.
Conceptual Knowledge	I want to know how and why PETs work.	I want to know how PETs are integrated into a system architecture.	I want to be able to create my own architectures involving PETs.	I want to be able to compare PETs based on their principal attributes.	I want to decide on which PET would be most suitable in a given system environment.	I want to create meta-models for PETs.
Procedural Knowledge	I want to identify use cases for applying PETs.	I want to be able to explain how different PETs are implemented.	I want to be able to implement PETs in a system environment.	I want to compare different ways to implement PETs.	I want to decide on the best way to implement a PET in a given situation.	I want to contribute to the development of new PETs.
Metacognitive Knowledge	I want to know the limitations of PETs.	I want to identify the limitations of a given PET implementation.	I want to be able to give strategies for optimizing the implementation of PETs.	I want to compare PET implementations based on their effectiveness.	I want to evaluate PET implementations and develop recommendations.	I want to find new use cases to which PETs could be applied.

Table 4
Survey answers per privacy role category

(a) Technical Experts

	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge	15	14	8	3	2	0
Conceptual Knowledge	15	12	11	6	4	0
Procedural Knowledge	15	13	11	7	7	0
Metacognitive Knowledge	15	12	8	2	2	2

(b) Legal Experts

	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge	6	4	4	4	4	2
Conceptual Knowledge	6	3	2	2	2	0
Procedural Knowledge	6	3	2	2	2	0
Metacognitive Knowledge	6	4	3	3	1	1

(c) Go-Betweens

	Remember	Understand	Apply	Analyze	Evaluate	Create
Factual Knowledge	3	3	1	1	1	1
Conceptual Knowledge	3	3	2	2	2	1
Procedural Knowledge	3	2	0	0	0	0
Metacognitive Knowledge	3	2	2	2	1	1

5. Role-Specific Educational Needs

To evaluate the relevance of the presented in Table 3 learning goals for different roles, we conducted a survey with practitioners. In designing the survey, we first ensured that the role of each respondent was captured. Next, the statements of each separate knowledge level were presented, and the respondent was prompted to select which statement was most relevant to the task of their specific privacy role. In addition, the respondent was informed that the statements followed a hierarchical order, meaning that selecting a more advanced cognitive process included all the previous ones as relevant. For example, in the Factual Knowledge category, choosing "I want to know the various use cases for PETs" implies that "I want to know what different PETs exist" also applies.

The role-specific insights are presented in Table 4 which separates the results based on the reported role. Table 4 utilizes a heat map to illustrate the frequency by which a particular option was chosen. Thus, the number displayed in each cell represents the aggregated number of responses that the corresponding option received, considering the previously introduced hierarchical setup.

As can be seen from Table 4, roles from the three different privacy role categories possess different learning goals, which is made particularly salient by our utilization of Bloom's Revised Taxonomy. Table 4b suggests that legal experts in the privacy compliance process would be most concerned with obtaining factual knowledge about PETs. This is plausible, as legal experts would not be involved in the implementation of PETs, but rather must be knowledgeable on the topic in general, i.e., know the facts. In Table 4a, a clear preference from technical experts towards factual and procedural knowledge can be observed. Thus, these experts must not only be cognizant of the facts, but also be skilled in the procedural know-how required for the implementation of PETs. Another interesting finding arrives with an analysis of the learning goals of Go-Between roles, whose preferences seemingly reside distinctly in conceptual knowledge. Looking to Table 3 for an explanation, one can see that conceptual knowledge truly lies on the border between factual and procedural knowledge, in the way that factual knowledge becomes important more from an IT architecture and policy point of view, rather than pure implementation. Indeed, members of

the Go-Between category do exist to bridge this gap, serving as a crucial link between legal mandate and technical specification.

6. Conclusion and Outlook

In this research in progress, we explore the educational needs of privacy professionals with respect to learning about Privacy-Enhancing Technologies. Under the framework of Bloom's Revised Taxonomy, we subdivide PET education into learning goals based on six cognitive processes and four knowledge levels. Moreover, we probe the relevance of each of these categories with different subgroups of privacy professionals: technical and legal experts, as well as Go-Betweens. The results of the survey provide insights into differing educational needs governed by the requirements of each role.

The practical relevance of this work is grounded in the underlying complexities of state-of-the-art PETs, which, without the necessary expertise, can hinder their adoption, calling for focused educational efforts to foster the development of such expertise. Looking forward, we plan not only to continue working on making knowledge on PETs open, accessible, and understandable, but also to do so in a way that considers the expertise of the learner. Our next steps include the creation of learning material on PETs, the validation of such material, and the deployment of an e-learning platform to encapsulate the learning content. In the creation of learning material, the findings presented in this work will be integral to tailoring the learning experience to different professional backgrounds with specific learning needs. The e-learning platform will provide the opportunity for collaboration with industry partners, further closing the gap between academia and industry on the topic of Privacy-Enhancing Technologies.

Acknowledgements

This work has been supported by the German Federal Ministry of Education and Research (BMBF) Software Campus grant LACE 01IS17049.

References

- [1] J. Wolff and N. Atallah, Early GDPR Penalties: Analysis of Implementation and Fines Through May 2020, *Journal of Information Policy*, vol. 11, no. 1, pp. 63–103, Jan. 2021.
- [2] M. Goddard, The EU General Data Protection Regulation (GDPR): European Regulation that has a Global Impact, *International Journal of Market Research*, vol. 59, no. 6, pp. 703–705, Nov. 2017.
- [3] O. Klymenko, S. Meisenbacher, and F. Matthes, Identifying Practical Challenges in the Implementation of Technical Measures for Data Privacy Compliance, *AMCIS 2023 Proceedings*, 2, 2023.
- [4] M. Hansen, J.-H. Hoepman, M. Jensen, and S. Schiffner, Readiness analysis for the adoption and evolution of privacy enhancing technologies: methodology, pilot assessment, and continuity plan, *Technical report: ENISA*, 2015.
- [5] O. Klymenko, O. Kosenkov, S. Meisenbacher, P. Elahidoost, D. Mendez, and F. Matthes, Understanding the Implementation of Technical Measures in the Process of Data Privacy Compliance: A Qualitative Study, in *Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement*, 2022, pp. 261–271.
- [6] S. Gürses and J. M. Del Alamo, Privacy engineering: Shaping an emerging field of research and practice, *IEEE Security & Privacy*, vol. 14, no. 2, pp. 40–46, 2016.
- [7] S. Fischer-Hübner and H. Lindskog, Teaching privacy-enhancing technologies, in *Proceedings of the IFIP WG 11.8 2nd World Conference on Information Security Education*, 2001, pp. 1–17.
- [8] D. R. Krathwohl, A Revision of Bloom's Taxonomy: An Overview, *Theory Into Practice*, vol. 41, no. 4, pp. 212–218, 2002.

- [9] G. W. Van Blarckom, J. J. Borking, and J. G. E. Olk, Handbook of privacy and privacy-enhancing technologies, Privacy Incorporated Software Agent (PISA) Consortium, The Hague, vol. 198, p. 14, 2003.
- [10] Information Commissioner's Office, Privacy-enhancing technologies (PETs), 2023.
- [11] The Royal Society, From privacy to partnership: the role of Privacy Enhancing Technologies in data governance and collaborative analysis, 2023.
- [12] United Nations, The United Nations Guide on Privacy-Enhancing Technologies for Official Statistics, 2023.
- [13] V. Braun and V. Clarke, Using thematic analysis in psychology, *Qualitative research in psychology*, vol. 3, no. 2, pp. 77–101, 2006.
- [14] M. Wiesche, M. C. Jurisch, P. W. Yetton, and H. Krcmar, Grounded theory methodology in information systems research, *MIS quarterly*, vol. 41, no. 3, pp. 685-A9, 2017.
- [15] L. W. Anderson and D. R. Krathwohl, *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. Longman, 2001.
- [16] C. Servin, C. Tang, M. Geissler, M. Stange, and C. Tucker, Enhanced Verbs for Bloom's Taxonomy with Focus on Computing and Technical Areas, in *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, Virtual Event, USA, 2021, p. 1270.

Nudging Towards Compliance? Assessing the Impact of Nudging Strategies on Information Security Policy Adherence

Theresa Pfaff ^{1,2}

¹ University of Goettingen, Goettingen, Germany

² University of Paderborn, Paderborn, Germany

Abstract

Data breaches pose a significant economic risk to companies in their daily business. To mitigate this risk, organizations implement information security policies (ISPs) to guide their employee's behavior. However, employees often fail to comply with these policies. To address this issue and promote desired behavior, the concept of nudging has emerged as a potential strategy. By leveraging insights from the dual-process theory, which recognizes two distinct cognitive systems involved in decision-making, this ongoing research aims to explore the effectiveness of nudging strategies through an online experiment. Specifically, it investigates whether information security policy messages can nudge employees towards adopting more secure behaviors by targeting their intuitive responses (System 1) or invoking critical thinking (System 2). This research seeks to advance our understanding of behavioral interventions in the context of information security and has the potential to provide valuable insights for designing effective strategies to promote ISP compliance.

Keywords

ISP Compliance, Digital Nudging Strategies, Dual-Process Theory

1. Introduction


One of the most prominent threats to organizational information assets comes from employees who have regular access to these resources [1, 2]. To mitigate the risks posed by insider threats, organizations adapt their risk management by implementing ISPs as a crucial instrument to reduce vulnerabilities and guide employee behavior. ISPs are a documented set of rules and guidelines that outline how an organization protects its sensitive information and manages information security risks [3]. It is a necessary tool as organizations face significant risks from cyberattacks and data breaches targeting their internal information, resulting in substantial costs for the affected company [4].


Research indicates that individuals often exhibit inappropriate and insecure behavior because they often prioritize convenience over adhering to information security policies [5]. Prior studies have examined various factors influencing employee compliance from a rational standpoint. Recent studies looked at employee's behavior by focusing on costs and benefits of compliance based on rational choice theory [6], threat and coping assessment by utilizing protection motivation theory [7, 8], and exerting pressure by using general deterrence theory including the assessment of potential sanctions [9, 10, 11]. While these studies provide valuable insights, they often focus on factors to explain certain behavior or to understand employees decision-making processes in a work environment. For example, most studies in this field specifically focus on employees' attitudes, knowledge, and intentions towards ISP compliance. They, furthermore, often explore the role of individual factors, such as awareness, perception of risk, and organizational support, in order to make the black-box of humans' decision-making-processes more transparent [12, 13]. While research has provided valuable insights into the factors that

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ theresa.pfaff@uni-goettingen.de; tpfaff@mail.uni-paderborn.de (T. Pfaff)

 0009-0005-1263-1760 (T. Pfaff)

 © 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

contribute to understanding security behavior, the question how to get employees towards the desired behavior has not been investigated yet. Deterrence measures for example, like punishment may effectively force employees towards the desired outcome [11]. Still, they also provoke a work environment built on fear and dissatisfaction. Surprisingly and to the best of our knowledge, no study has thought about the concept of nudging in this context i.e., how to strategically nudge employees towards the desired behavior. A recent study by [14] has shown that different types of employees react differently in their compliance behavior to certain deterrents. Therefore, it is reasonable to assume that this could also be the case with different nudge messages addressing different systems. In this context, nudging refers to the use of certain design elements in a user interface to influence users' choices while using IS [15].

At the same time, most studies only focus on factors affecting employee compliance on both a rational and deliberated level. Consequently, there is a need to delve deeper into the less-explored dimensions of employee compliance, also considering non-rational and automated aspects that may influence their behavior. This research aims to fill this gap by investigating the dual-process nature of employees' cognitive responses to information security nudge messages. Therefore, this study seeks to answer the research question (RQ):

RQ: How can information security policy (ISP) messages nudge employees towards a more compliant behavior?

The aim of this research is to investigate whether and how information security policy messages can effectively nudge employees towards enhanced compliance with ISPs as this is the desired behavior from an organizational perspective. The study further seeks to uncover the mechanisms through which these policy messages influence employees' cognitive processes, specifically their gut reactions (System 1) and/or critical thinking (System 2). By unraveling the impact of information security policy messages on employees' cognitive processes, the research aims to contribute to the understanding of behavioral interventions in the context of information security and provide insights for designing more effective strategies to promote ISP compliance. The aim of this research in progress paper is to present a status quo of current undertakings and to provide an outlook on further actions.

2. Theoretical Background

2.1. Dual-Process Theory

First The dual-process theory posits that human cognition and decision-making involve two distinct cognitive processes: System 1 and System 2. System 1 thinking is automatic, intuitive, and fast, driven by heuristics and immediate emotional responses. On the other hand, System 2 thinking is reflective, deliberate, and analytical, involving conscious reasoning and cognitive effort [16, 17]. In the context of investigating information security policy compliance, the dual-process theory provides a suitable theoretical lens for several reasons. In order to understand employees' compliance behavior, it requires examining both, their automatic, intuitive responses (System 1) and their reflective, deliberative processes (System 2). By considering the interplay between these cognitive processes, insights into the factors influencing employees' decision-making can be gained. For example, [18] found that presenting fact-checking results in a combined approach targeting both systems, automatic cognition via symbols and deliberate cognition via text phrases, was twice as effective in detecting fake news compared to settings where only one system was primarily addressed. Additionally, the theory helps to explain why employees may exhibit inconsistent compliance behavior. System 1 responses, driven by heuristics and emotions, can lead to impulsive or careless actions that deviate from established policies [19]. System 2 thinking, on the other hand, allows employees to engage in conscious reasoning and critically evaluate the implications of their behavior in terms of information security. The dual-process theory also highlights the potential conflicts and trade-offs between System 1 and System 2 processes [20, 21]. Employees may face cognitive biases, such as cognitive dissonance or anchoring, that influence their decision-making and adherence to security policies.

Understanding these conflicts can provide valuable insights for designing effective nudging strategies that target both automatic and reflective cognitive processes.

Overall, the dual-process theory provides a comprehensive framework for examining the cognitive mechanisms underlying employees' compliance behavior and offers guidance for designing interventions that effectively promote information security policy compliance. paragraph in every section does not have first-line indent. Use only styles embedded in the document.

2.2. Digital Nudging towards Desired Behavior

Digital nudging refers to the strategic use of subtle and non-intrusive digital interventions aimed at guiding individuals' decision-making and influencing their behavior towards desired outcomes [15]. Rooted in behavioral economics and psychology, digital nudging leverages principles of choice architecture to shape decisions without resorting to strict regulations or mandates. In the context of employees' compliance with ISPs, investigating nudging strategies becomes imperative due to the persistent challenge of motivating employees to adhere to established security protocols. Traditional approaches, such as training programs and enforcement measures, often fall short in effectively modifying employees' behavior. By exploring the potential of digital nudging techniques, organizations can harness the power of choice architecture to nudge employees towards more secure behaviors.

This research endeavors to examine the efficacy of nudging strategies in the realm of ISP compliance, aiming to provide valuable insights into the design of interventions that align with employees' decision-making processes, thereby fostering a culture of enhanced information security while respecting individuals' autonomy and decision-making agency. Nudging strategies are applied in various contexts, such as public health decisions, consumer behavior, or tax compliance [15, 22, 23]. While there is an upcoming trend of examining nudging in privacy and security context, research primarily focuses on privacy settings, password creation or phishing detection [24, 25, 26]. However, recent literature claims to further extend the design of nudges to other scenarios in cybersecurity, such as protection of data [26].

3. Hypotheses Development and Research Model

The goal of an intervention aiming to influence cognitive functions in System 1 is to provide an intuitively clear stimulus, according to [18]. Simple visual signs can be understood fast and with less cognitive effort [21] making it a suitable nudge strategy, triggering heuristic and immediate responses. Contrary to this, textual detailed information will more likely trigger System 2 as it takes more time and effort to process the given information. Understanding text arguments and connecting them to prior knowledge requires deliberate attention to detail, which is usually part of System 2 cognition [17]. In their study, [24] revealed that a security nudge text-message can increase users security behavior. Especially messages emphasizing the threat and the corresponding coping behavior were most effective. Therefore, it is expected:

H1a: Employees' ISP compliance behavior with a System 1 nudge strategy is enhanced when compared to decision settings in which no nudge is applied.

H1b: Employees' ISP compliance behavior with a System 2 nudge strategy is enhanced when compared to decision settings in which no nudge is applied.

However, system 1 and system 2 cannot be strictly separated as both systems are considered rather complements than substitutes [18]. An intervention that combines the two theories will likely have a greater effect than the two single-interventions if they are both, primarily, acting through one theoretical route. If both single interventions primarily act through one theoretical route, then an intervention that combines the two strategies and triggers both systems simultaneously will likely have a better effect than the single-interventions [15]. Therefore, it is stated:

H2: The combination of both nudging strategies is more effective than no nudge or a single nudge applied.

Information security policy compliance is more likely to occur if employees believe their managers, IT personnel, or peers expect them to comply [12]. However, work environments are dynamic environments with arising situational characteristics such as demanding colleagues and finding workarounds [27]. Conversely, if peers, IT personnel, or managers themselves do not adhere to policies this, then employees might adapt this behavior. Moreover, if these peers put colleagues into demanding situations this, can result in peer pressure which is defined as influencing or urging individuals to do something, regardless of whether they personally want to or not [28]. It can be argued that peers not following ISPs will cause other employees to break the rules and diminish the effect of nudge messages towards ISP compliance, leading to the following hypotheses:

H3: The perceived peer pressure from colleagues to deviate from the ISP will negatively impact employees' ISP compliance behavior.

H4a: The perceived peer pressure from colleagues to deviate from the ISP will weaken the impact of the System 1 nudge strategy on employees' ISP compliance.

H4b: The perceived peer pressure from colleagues to deviate from the ISP will weaken the impact of the System 2 nudge strategy on employees' ISP compliance.

H5: The perceived peer pressure from colleagues to deviate from the ISP will weaken the impact of the combined nudge strategy on employees' ISP compliance.

Figure 1 presents the proposed research model.

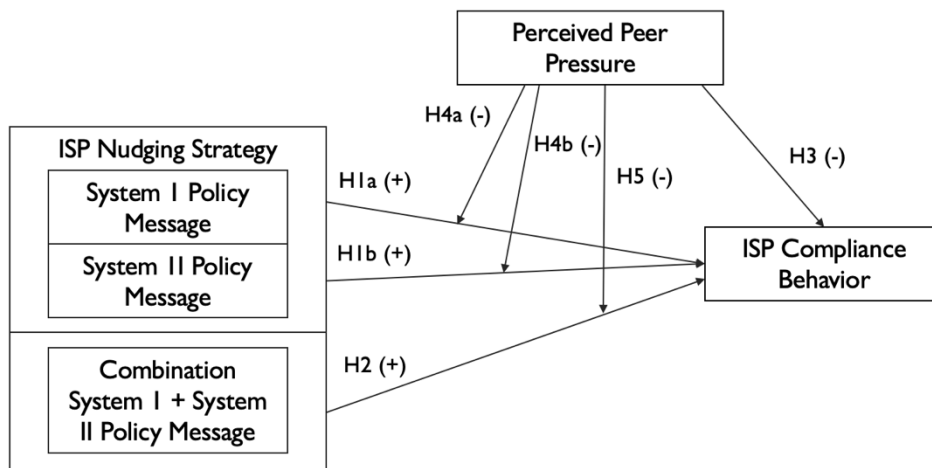


Figure 1: Research Model

4. Methodology

To test the proposed model, a four-condition between-subjects design will be applied in an online experiment where ISP compliance will be measured through an in-basket task. Participants will be required to respond to incoming mails from fictive colleagues, assessing their compliance behavior. An in-basket task with emails provides a realistic simulation of employees' work environment, allowing researchers to assess compliance behavior in authentic scenarios [29]. Emails, being a common communication mode in organizations, offer a relevant context for measuring ISP compliance. The design for this email task will follow [13], [30], [31]. For evaluation, a binary coding scheme will be used, assigning a value of "0" for non-compliance and "1" for compliance. The design of the ISP nudge messages is based on the approaches of [18], [32],

[33] and will appear right before participants start with their task. Perceived peer pressure will be assessed using a 7-point Likert scale ranging from 1="strongly disagree" to 7="strongly agree". The specific items, as well as the complete in-basket task, are currently under development and will be presented in a future research paper. A potential challenge in this study is the power of the nudge message, which needs to be strong enough to interfere across multiple mails. Therefore, a pilot study will be carried out first. To conduct data analysis and test the proposed research model, H1a/b will be evaluated using unpaired t-tests while H2 will be assessed with a one-way ANOVA. Hypotheses 3 to 5 will be tested using structural equation modelling (SEM) with the software SmartPLS.

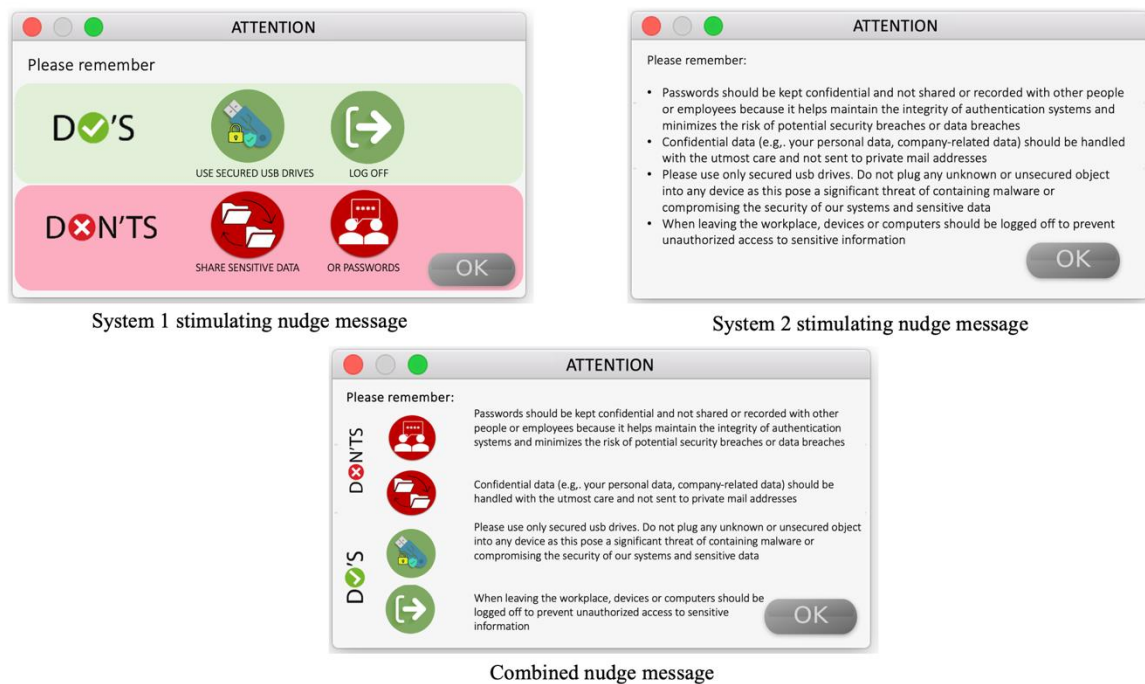


Figure 2: Nudge Message Designs

5. Conclusion

In conclusion, this study aims to investigate the effectiveness of different nudging strategies in enhancing employees' ISP compliance behavior. Building upon the dual-process theory, which suggests that individuals' decision-making can be influenced by both intuitive (System 1) and reflective (System 2) processes, the study explores the impact of System 1 and System 2 digital nudge messages on employees' ISP compliance behavior. A potential challenge in this study is the power of the nudge message, which needs to be strong enough to interfere across multiple mails. By examining the individual and combined effects of visual and textual nudges, the study seeks to provide insights into the mechanisms through which these nudges influence employees' cognitive processes. The findings of this study will contribute to the understanding of behavioral interventions in the context of information security. Ultimately, the study aims to advance knowledge in the field and provide practical recommendations for organizations seeking to improve their employees' adherence to information security policies.

References

- [1] M. Warkentin and R. Willison, "Behavioral and policy issues in information systems security: The insider threat," *Eur. J. Inf. Syst.*, vol. 18, no. 2, pp. 101–105, 2009, doi: 10.1057/ejis.2009.12.
- [2] R. Willison and M. Warkentin, "Beyond deterrence: An expanded view of employee computer abuse," *MIS Q. Manag. Inf. Syst.*, vol. 37, no. 1, pp. 1–20, 2013, doi: 10.25300/MISQ/2013/37.1.01.
- [3] D. Ormond, M. Warkentin, and R. E. Crossler, "Integrating cognition with an affective lens to better understand information security policy compliance," *J. Assoc. Inf. Syst.*, vol. 20, no. 12, pp. 1794–1843, 2019, doi: 10.17705/1jais.00586.
- [4] H. Cavusoglu, B. Mishra, and S. Raghunathan, "A model for evaluating IT security investments," *Commun. ACM*, vol. 47, no. 7, pp. 87–92, 2004, doi: 10.1145/1005817.1005828.
- [5] J. D'Arcy and P. B. Lowry, "Cognitive-affective drivers of employees' daily compliance with information security policies: A multilevel, longitudinal study," *Inf. Syst. J.*, vol. 29, no. 1, pp. 43–69, 2019, doi: 10.1111/isj.12173.
- [6] H. Li, J. Zhang, and R. Sarathy, "Understanding compliance with internet use policy from the perspective of rational choice theory," *Decis. Support Syst.*, vol. 48, no. 4, pp. 635–645, 2010, doi: 10.1016/j.dss.2009.12.005.
- [7] G. D. Moody, M. Siponen, and S. Pahnla, "Toward a unified model of information security policy compliance," *MIS Q. Manag. Inf. Syst.*, vol. 42, no. 1, pp. 285–311, 2018, doi: 10.25300/MISQ/2018/13853.
- [8] K. Masuch, S. Hengstler, S. Trang, and A. B. Brendel, "Replication Research of Moody, Siponen, and Pahnla's Unified Model of Information Security Policy Compliance," *AIS Trans. Replication Res.*, vol. 6, no. 13, pp. 1–16, 2020, doi: 10.17705/1attr.00056.
- [9] J. D'Arcy and T. Herath, "A review and analysis of deterrence theory in the IS security literature: Making sense of the disparate findings," *Eur. J. Inf. Syst.*, vol. 20, no. 6, pp. 643–658, 2011, doi: 10.1057/ejis.2011.23.
- [10] J. D'Arcy, T. Herath, and M. K. Shoss, "Understanding Employee Responses to Stressful Information Security Requirements: A Coping Perspective," *J. Manag. Inf. Syst.*, vol. 31, no. 2, pp. 285–318, 2014, doi: 10.2753/MIS0742-1222310210.
- [11] S. Trang and B. Brendel, "A Meta-Analysis of Deterrence Theory in Information Security Policy Compliance Research," *Inf. Syst. Front.*, vol. 21, no. 6, pp. 1265–1284, 2019, doi: 10.1007/s10796-019-09956-4.
- [12] B. Bulgurcu, H. Cavusoglu, and I. Benbasat, "Information Security Policy Compliance: An Empirical Study of Rationality-Based Beliefs and Information Security Awareness," *MIS Q. Manag. Inf. Syst.*, vol. 34, no. 3, pp. 523–548, 2010.
- [13] L. Jaeger and A. Eckhardt, "When colleagues fail: Examining the role of information security awareness on extra-role security behaviors," *26th Eur. Conf. Inf. Syst. Beyond Digit. - Facet. Socio-Technical Chang. ECIS 2018*, 2018.
- [14] S. Hengstler, S. Kuehnel, K. Masuch, I. Nastjuk, and S. Trang, "Should I Really do That? Using Quantile Regression to Examine the Impact of Sanctions on Information Security Policy Compliance Behavior," *Comput. Secur.*, vol. 133, p. 103370, 2023, doi: 10.1016/j.cose.2023.103370.
- [15] M. Mirbabaie, J. Marx, and J. Germies, "Conscious Commerce-Digital Nudging and Sustainable E-commerce Purchase Decisions," *Australas. Conf. Inf. Syst.*, pp. 1–11, 2021.
- [16] K. E. Stanovich and R. F. West, "Individual differences in reasoning: Implications for the rationality debate?," *Behav. Brain Sci.*, vol. 26, no. 4, p. 527, 2003, doi: 10.1017/S0140525X03210116.
- [17] D. Kahneman, "A Perspective on Judgment and Choice: Mapping Bounded Rationality," *Am. Psychol.*, vol. 58, no. 9, pp. 697–720, 2003, doi: 10.1037/0003-066X.58.9.697.

- [18] P. L. Moravec, A. Kim, and A. R. Dennis, "Appealing to sense and sensibility: System 1 and system 2 interventions for fake news on social media," *Inf. Syst. Res.*, vol. 31, no. 3, pp. 987–1006, 2020, doi: 10.1287/ISRE.2020.0927.
- [19] J. S. B. T. Evans and K. E. Stanovich, "Dual-Process Theories of Higher Cognition: Advancing the Debate," *Perspect. Psychol. Sci.*, vol. 8, no. 3, pp. 223–241, 2013, doi: 10.1177/1745691612460685.
- [20] J. S. B. T. Evans and J. Curtis-Holmes, "Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning," *Think. Reason.*, vol. 11, no. 4, pp. 382–389, 2005, doi: 10.1080/13546780542000005.
- [21] J. S. B. T. Evans, "Dual-processing accounts of reasoning, judgment, and social cognition," *Annu. Rev. Psychol.*, vol. 59, pp. 255–278, 2008, doi: 10.1146/annurev.psych.59.103006.093629.
- [22] J. Wisdom, J. S. Downs, and G. Loewenstein, "Promoting Healthy Choices : Information versus Convenience Author (s): Jessica Wisdom , Julie S . Downs and George Loewenstein Published by : American Economic Association Stable URL : <https://www.jstor.org/stable/25760210> REFERENCES Linked references a," *Am. Econ. J. Appl. Econ.*, vol. 2, no. 2, pp. 164–178, 2010.
- [23] A. Antinyan and Z. Asatryan, "Nudging for Tax Compliance: A Meta-Analysis," *SSRN Electron. J.*, no. 8500, 2021, doi: 10.2139/ssrn.3680357.
- [24] R. van Bavel, N. Rodríguez-Priego, J. Vila, and P. Briggs, "Using protection motivation theory in the design of nudges to improve online security behavior," *Int. J. Hum. Comput. Stud.*, vol. 123, no. September 2018, pp. 29–39, 2019, doi: 10.1016/j.ijhcs.2018.11.003.
- [25] V. Zimmermann and K. Renaud, "The nudge puzzle: Matching nudge interventions to cybersecurity decisions," *ACM Trans. Comput. Interact.*, vol. 28, no. 1, 2021, doi: 10.1145/3429888.
- [26] K. Hartwig and C. Reuter, "Nudge or restraint: How do people assess nudging in cybersecurity - A representative study in germany," *ACM Int. Conf. Proceeding Ser.*, pp. 141–150, 2021, doi: 10.1145/3481357.3481514.
- [27] I. Kirlappos, S. Parkin, and M. A. Sasse, "Learning from 'Shadow Security:' Why Understanding Non-Compliant Behaviors Provides the Basis for Effective Security," *USEC'14 Work. Usable Secur.*, no. February, 2014, doi: 10.14722/usec.2014.23007.
- [28] D. R. Clasen and B. B. Brown, "The multidimensionality of peer pressure in adolescence," *J. Youth Adolesc.*, vol. 14, no. 6, pp. 451–468, 1985, doi: 10.1007/BF02139520.
- [29] D. S. Kiker and S. J. Motowidlo, "Main and interaction effects of task and contextual performance on supervisory reward decisions," *J. Appl. Psychol.*, vol. 84, no. 4, pp. 602–609, 1999, doi: 10.1037/0021-9010.84.4.602.
- [30] M. Siponen and A. Vance, "Neutralization: New Insights into the Problem of Employee Information Systems Security Policy Violations," *MIS Q. Manag. Inf. Syst.*, vol. 34, no. 3, pp. 487–502, 2010.
- [31] S. Trang and I. Nastjuk, "Examining the role of stress and information security policy design in information security compliance behaviour: An experimental study of in-task behaviour," *Comput. Secur.*, vol. 104, p. 102222, 2021, doi: 10.1016/j.cose.2021.102222.
- [32] C. Schneider, M. Weinmann, and J. Vom Brocke, "Digital nudging: Guiding online user choices through interface design Designers can create designs that nudge users toward the most desirable option," *Commun. ACM*, vol. 61, no. 7, pp. 67–73, 2018, doi: 10.1145/3213765.
- [33] T. Mirsch, C. Lehrer, and R. Jung, "Making digital nudging applicable: The digital nudge design method," *Int. Conf. Inf. Syst. 2018, ICIS 2018*, no. 2009, pp. 1–16, 2018.

How to Foster Compliance in Non-Integrated IT-Landscapes? The Case of Manual Medical Data Transfers

Gilbert Georg Hövel¹ and Tizian Matschak^{1,2}

¹ Paderborn University, Warburger Straße 100, Paderborn, 33098, Germany

² University of Goettingen, Platz der Göttinger Sieben 5, Goettingen, 37073, Germany

Abstract

Due to the slow pace of digital transformation in many industries, IT-landscapes are still often non-integrated. Therefore, in industries with non-integrated IT-landscapes professionals still transfer data manually. One prominent example is the healthcare sector. Medical professionals often need to transfer medication data between different Health Information Systems (HIS) manually. Errors that occur during this manual procedure often go unnoticed and can have far-reaching health-consequences for patients. Based on the Deterrence Theory, we plan to examine how different formal sanction mechanisms are related to various types of medication errors. In doing so, we aim to demonstrate how sanction mechanisms can foster compliance in non-integrated IT-landscapes. In investigating medication errors from an organizational lens, we aim to extend current research on medication errors.

Keywords

Compliance in Healthcare, Formal Sanction Mechanisms in Digital Health, Medication Errors

1. Introduction

In many industries digital transformation is progressing slowly. As result, a significant share of Information Systems (IS) is still non-integrated. This means that these IS are not interoperable and data often cannot be exchanged in a standardized way. As result, data needs to be transferred between different IS manually, making the process of data transfers more prone to errors [1]. One prominent example for an industry with many non-integrated IT-landscapes is the healthcare sector. Although there is a multitude of digitalization initiatives aiming to integrate the healthcare IT-landscapes, medication data still need to be transferred between different Health Information Systems (HIS) manually. The manual medication data transfers often lead to errors [2].

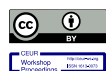
In healthcare, medication errors are one of the most frequently occurring error type [3]. According to the World Health Organization (WHO) 10% of hospitalizations are a direct result of medication errors. Furthermore, the WHO estimates the costs associated with medication errors to exceed 40 billion USD each year globally [4]. Medication errors can occur in the manual procedures of prescribing, transcribing, dispensing, administering, and monitoring of medications (e.g., missing data) [5]. Medication errors can result in serious health consequences for patients and may even lead to a patient's death [5].

It can be assumed that in integrated IT-landscapes errors occur significantly less often, since the data can be transferred automatically [1,6]. However, as it will take some time until more IT-landscapes in healthcare are fully integrated and not all stakeholders may will be willing to integrate their systems, further research on the avoidance of medication errors is necessary.

Medication errors were identified as potential problems within Information Systems research years ago. However, in IS research, medication errors have only been investigated to a limited extent. Most studies in IS research investigate how Health Information Systems need to be designed to avoid medication errors (e.g., [7]). There is a rich body of literature on medication

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ gilbert.georg.hoewel@upb.de (G. G. Hövel); tizian.matschak@upb.de (T. Matschak)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
CEUR Workshop Proceedings (CEUR-WS.org)

errors investigating how to prevent medication errors in medical science and the field of medical informatics. Many of these studies conduct real-world interventions and investigate the phenomenon in retrospective examinations, for example by analyzing medical documentations [5,6,8]. This approach is often applied in IS research as well [5,6]. Thus, there is still a lack of research that investigates organizational mechanisms that can help to prevent medication errors.

Primarily, research shows that time constraints, interruption during the manual data transfer, and inattention are reasons for medical errors [8]. In addition, as manual data transfers are time-consuming, it can be assumed that errors occur because medical professionals want to save time and risk to transferring the data inaccurately. Generally, medical professionals are responsible for the correctness of the medication data when transferring it. Thus, errors in the data transfer can be considered as a medical professionals' non-compliance. As medication errors are rarely identified, the probability that this non-compliance will be detected is low [8]. Therefore, missing sanctions may foster medical professionals' non-compliance.

Compliance research has shown that organizational sanction mechanisms may help to avoid professionals' non-compliance [9]. In healthcare, organizational sanction mechanisms are for instance implemented by defining and reviewing clinical guidelines [6]. Violating the guidelines can be sanctioned by disciplinary actions. To this background, we aim to study how sanctions can be utilized to avoid medical errors. By doing so, we contribute to compliance research by linking sanction mechanisms to different forms of non-compliance. Furthermore, we extend the literature in the domain of digital health by presenting organizational mechanisms that can help to prevent medications errors. Our research offers valuable insights to define policies that can help to prevent medication errors and can be transferred to other areas with non-integrated IT-landscapes. Accordingly, this paper aims to answer the following research question:

RQ: How do organizational compliance mechanisms affect different kinds of data transfer errors in non-integrated IT-landscapes?

This research-in-progress paper introduces the identified research problem and outlines the planned research approach. The remainder of this paper is organized as follows: First, we provide an overview of the contextual background and the theoretical foundation. Second, we present our research model. Lastly, we outline our planned research design.

2. Contextual Background and Theoretical Foundation

2.1. Medication Data Transfers in Practice

In general, healthcare IT infrastructures involve different stakeholders such as primary care, hospitals, and health insurances. The healthcare sector faces the problem of many stakeholders operating their own IT systems which merely coexist. These HIS often store health data in different formats. Furthermore, many processes in healthcare are still paper-based. As mentioned before, that has the consequence that health data often cannot be exchanged in a standardized way [10].

In the healthcare sector, there is a multitude of digitalization initiatives such governmental initiatives that aim to allow patients to collect their health data in electronic health records (EHR) exist in many countries [10]. This for instance provides the opportunity to integrate the patients' medication data into HIS automatically. However, digital transformation is progressing slowly, and it will take some time until all healthcare stakeholders are integrated efficiently. For example, the rollout of the electronic health records in Germany started in 2021 and the rollout is still continuing [11]. This means, for instance, that medical data cannot be transferred digitally when patients are admitted to a hospital. Until EHR are rolled out completely, healthcare stakeholders are instructed to print out standardized medication plans in Germany [11] (see Table 1). Since not all healthcare stakeholders even have HIS, medication data still need to be transferred manually most of the time.

Table 1
Types of Medication Plans

Hand-written	Standardized-Printout																																																																																								
<p>Transl. lang. 250 1-1-1 + Nahrungsmittel 500g 1-1-1 Vitamin D3 HL150 3x 1-1-1 Tag 1x abends 1x abends Urtio Macrogol 1H 2-3x am Tag Cefuroxim 500mg 1x abends Panthenol 2x abends</p>	<div style="border: 1px solid black; padding: 5px;"> <p>Medikationsplan Für: Maxi Mustermann Geb. am: 01.01.1960</p> <p>Ausgedruckt von: Pharm Dr. Armin Hölzger Gabelweg, 6 Berlin Telefon: 030 2200 E-Mail: dr_armin_hoelzger_at_net Ausgedruckt am: 13.08.2020</p> <table border="1"> <thead> <tr> <th>Wirkstoff</th> <th>Handelsname</th> <th>Stärke</th> <th>Form</th> <th>Mo- gens</th> <th>Mittags</th> <th>Abends</th> <th>Zur Nacht</th> <th>Einheit</th> <th>Hinweise</th> <th>Grund</th> </tr> </thead> <tbody> <tr> <td>Bisoprololumal Harnschonband</td> <td>Bisoprolol plus 5/12,5-1A</td> <td>5,0 mg 12,5 mg</td> <td>Förntabl</td> <td>1</td> <td>0</td> <td>0</td> <td>0</td> <td>Stück</td> <td></td> <td>Bluthochdruck</td> </tr> <tr> <td>Simvastatin</td> <td>Simvastatin - 1A</td> <td>10 mg</td> <td>Förntabl</td> <td>0</td> <td>0</td> <td>1</td> <td>0</td> <td>Stück</td> <td></td> <td>Blutfette</td> </tr> <tr> <td>Mellin</td> <td>Mellin</td> <td>500 mg</td> <td>Förntabl</td> <td>1</td> <td>0</td> <td>1</td> <td>0</td> <td>Stück</td> <td>Während oder nach den Mahlzeiten</td> <td>Blutzucker</td> </tr> </tbody> </table> <p>Zu besonderen Zeiten anzuwendende Medikamente</p> <table border="1"> <thead> <tr> <th>Arbeits-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> </tr> </thead> <tbody> <tr> <td>Hypertonie</td> <td>Arbeits-</td> <td>1,92 mg</td> <td>Einzel-</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> <p>Selbstmedikation</p> <table border="1"> <thead> <tr> <th>Arbeits-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> <th>Einzel-</th> </tr> </thead> <tbody> <tr> <td>Rötchenentzündung (Moracolin)</td> <td>Roter Riss Extrakt Tabletten</td> <td>4,5 mg</td> <td>Kapseln</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table> </div>	Wirkstoff	Handelsname	Stärke	Form	Mo- gens	Mittags	Abends	Zur Nacht	Einheit	Hinweise	Grund	Bisoprololumal Harnschonband	Bisoprolol plus 5/12,5-1A	5,0 mg 12,5 mg	Förntabl	1	0	0	0	Stück		Bluthochdruck	Simvastatin	Simvastatin - 1A	10 mg	Förntabl	0	0	1	0	Stück		Blutfette	Mellin	Mellin	500 mg	Förntabl	1	0	1	0	Stück	Während oder nach den Mahlzeiten	Blutzucker	Arbeits-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Hypertonie	Arbeits-	1,92 mg	Einzel-								Arbeits-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Rötchenentzündung (Moracolin)	Roter Riss Extrakt Tabletten	4,5 mg	Kapseln							
Wirkstoff	Handelsname	Stärke	Form	Mo- gens	Mittags	Abends	Zur Nacht	Einheit	Hinweise	Grund																																																																															
Bisoprololumal Harnschonband	Bisoprolol plus 5/12,5-1A	5,0 mg 12,5 mg	Förntabl	1	0	0	0	Stück		Bluthochdruck																																																																															
Simvastatin	Simvastatin - 1A	10 mg	Förntabl	0	0	1	0	Stück		Blutfette																																																																															
Mellin	Mellin	500 mg	Förntabl	1	0	1	0	Stück	Während oder nach den Mahlzeiten	Blutzucker																																																																															
Arbeits-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-																																																																															
Hypertonie	Arbeits-	1,92 mg	Einzel-																																																																																						
Arbeits-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-	Einzel-																																																																															
Rötchenentzündung (Moracolin)	Roter Riss Extrakt Tabletten	4,5 mg	Kapseln																																																																																						
[11]	[12]																																																																																								

Medication data generally contains the following information: the names of the prescribed medications, information on the dose in which the medications are provided and, the frequency the patient receives the medications [8]. As the types of medication plans in Table 1 suggest, manual transfers of medication are accompanied by the risk of data being transferred incorrectly or incompletely. According to Callen et al. 2010 the following errors in manual medication data transfers can occur: *data is omitted, data is transferred inaccurately, and data is listed additionally* [8].

2.2. Research on Medication Errors

Research on errors in digital health distinguishes between interpretive and procedural errors [6]. Interpretive errors are based on the subjectivity of a decision [6]. An example for this are false diagnoses, as diseases are not always clearly identifiable. Procedural errors refer to deviations from norms and standards [6]. Since physicians are responsible for transferring the medication data correctly, medication errors are procedural errors.

To reduce procedural errors, corresponding literature suggests specifying procedural rules, observing and recording clinical actions, and reviewing medical professionals' compliance on a regular basis [6]. From this approach it becomes apparent that besides technical factors such as the design of HIS, it is also relevant to consider human, socio-technical, and organizational factors to prevent medication errors [14]. In line with that, studies on medication errors identified a wide range of causes which go beyond the design of HIS. Examples for human factors that cause medication errors are a lack of physical well-being and the resulting lack of concentration. A prominent socio-technical factor is physicians' missing attitude towards the use of HIS. From an organizational perspective, physicians often face heavy workloads which result in time pressure [8]. Design factors for instance refer to the structure and design of the HIS interfaces (e.g., [7]).

Although human and organizational factors are of particular interest to prevent medication errors, most of the existing research in the context of medical errors aims to avoid errors by improving the design of HIS. As one of the key approaches, corresponding literature explores the validation of the medical professionals' input and system notifications that display identified errors [7,15].

Since medication errors also depend on whether medical professionals even enter the data into the system, we argue that organizational mechanisms need to be defined in addition to implementing system notifications for incorrectly input content.

2.3. Deterrence Theory

Although medication errors can have far-reaching consequences for patients and are relevant, medication errors will rarely be identified [5,8]. Based on those circumstances, it can be assumed that medical professionals perceive the risk of errors being detected as low [7]. To ensure medical quality, healthcare organizations rely on formal compliance mechanism such defining clinical guidelines. To explain why these formal compliance mechanism work, research often draw on the Deterrence Theory (DT) [9,16].

Following the DT, people compare the probable costs and benefits of an undesired behavior. The DT originates from the field of criminology and aims to explain how people decide whether they commit a criminal act or not [16]. In this manner, the DT argues that the lower the external punishment, the more likely an individual decide for commit the criminal act [17].

The DT assumes that the expected punishment is influenced by the sanction certainty, severity, and celerity. The perceived certainty describes how likely an individual belief a potential sanction occurs. The perceived severity determines how strong the potential sanction is expected to be. The perceived celerity refers to the individuals' assessment how fast the sanction is given [18].

3. Hypotheses and Research Model

We aim to study the influence of medical professionals' perceived sanction severity, certainty, and celerity on human errors in manual medical data transfer. In particular, we aim to investigate whether perceived sanction severity, certainty, and celerity relate to different kinds of human errors in the medical data transfer.

Based on the three formal sanction mechanisms from the DT and the three error types mentioned in section two, we propose a research model with nine hypotheses (see Figure 1).

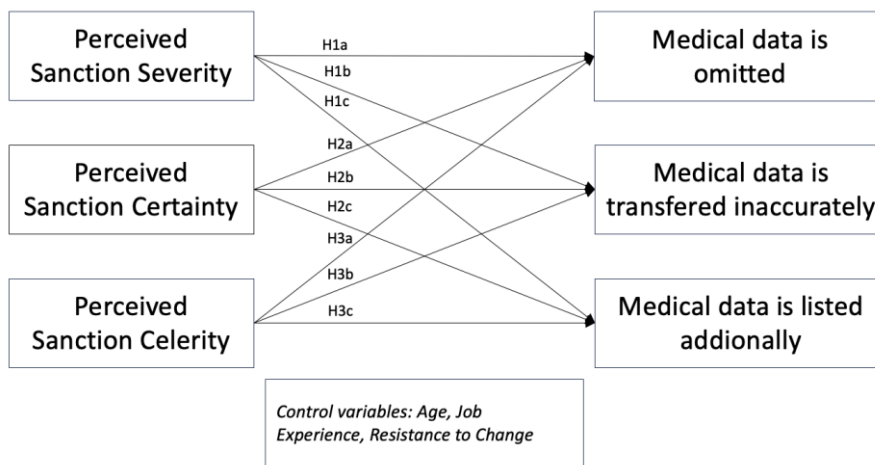


Figure 1: Research Model

Compared to other fields of compliance research, non-compliance can have far-reaching consequences for medical professionals. Medical professionals can be sanctioned internally (e.g., a hospital or the department of a hospital) but also externally (e.g., responsible authorities). In certain cases, medical professionals even risk losing their professional license. We therefore assume that medical professionals weigh the potential sanctions and benefits, such as time saved, in the process of transferring data.

In the context of manual medical data transfers, the severity of sanction describes the perceived impact a medical professional believes the potential sanction will have. Corresponding literature shows that a high perceived sanction severity discourages employees from non-compliant behaviors [9]. In the context of medication errors, it can be assumed that this mechanism is particularly effective as formal sanctions can have serious consequences for medical professionals (see Section 1). Hence, we derive the following hypotheses:

- H1a: *The higher the perceived sanction severity, the less medical data is omitted.*
 H1b: *The higher the perceived sanction severity, the less medical data is transferred inaccurately.*
 H1c: *The higher the perceived sanction severity, the less medical data is listed additionally.*

Perceived sanction certainty refers to the degree of likelihood a medical professional believes a sanction holds. Recent studies reveal that perceived sanction certainty is negatively associated with non-compliance, as the high likelihood of being detected increases the costs of non-compliant behaviors increase (e.g., [19,20]). As mentioned before, medication errors often remain unnoticed and medical professionals therefore assess the risk of being detected as low (see Section 1). Thus, increasing the sanction certainty seems to be a promising mechanism to avoid medication errors. Thus, we formulate the following hypotheses:

- H2a: *The higher the perceived sanction certainty, the less medical data is omitted.*
 H2b: *The higher the perceived sanction certainty, the less medical data is transferred inaccurately.*
 H2c: *The higher the perceived sanction certainty, the less medical data is listed additionally.*

Perceived sanction celerity relates to the period of time between the occurrence of the medication error and the sanction being pronounced. Studies found that swift sanctions affect employees' compliance positively since the sanction costs are decreasing with the time [9]. As outlined earlier, one important approach to avoid medication errors is validating the medication data input (see Section 2.2). In doing so, errors are identified immediately. Hence, medication errors can potentially be avoided by identifying and sanctioning these errors shortly after they appeared. Thus, we posit:

- H3a: *The higher the perceived celerity, the less medical data is omitted.*
 H3b: *The higher the perceived celerity, the less medical data is transferred inaccurately.*
 H3c: *The higher the perceived sanction celerity, the less medical data is listed additionally.*

4. Research Design and Method

4.1. Data Collection

To test the hypotheses, we plan to conduct an online experiment with medical professionals in a between-subject design. In the experiment, a manual medical data transfer from a medication plan to a HIS is simulated. The target participants are physicians and nurses because they are commonly involved in the manual transfer of medical data. Most importantly, physicians and nurses are able to assess the potential sanctions that result from human errors in the medical data transfer.

The data collection procedure is as follows. First, each participant receives a short introduction with explanations on the task. The task will be to enter medical data from a medication plan to an online formular within a given time. To provide a realistic scenario, the online formular includes key design-elements of a HIS. Each medication plan contains six prescribed medications. For each of these medications, the participants are advised to transfer the name of the medication, the dosage, and the frequency of use. After the task is performed, the participants will be asked to complete a questionnaire which contains the sanction mechanism constructs of the DT.

To manipulate the three sanction mechanisms from the DT, the experiment is structured in a 3x2 design (see Table 2). For each mechanism two scenarios (low and high) are defined through different representations of policy elements. The policy elements will be represented in the formular. Thereby, we rely on the suggestions of corresponding literature to review medical

professionals' compliance on a regular basis [6]. Furthermore, we bring our experiment in line with IS research on medication which mainly focuses on the interface of HIS (see Section 2.2).

Table 2
Experimental Design

Sanction	Scenario	Head 3
Severity	Low	Indication that in case an error is identified, the input need to be corrected
	High	Indication that in case an error is identified, the medical professional receives a warning by regulatory authorities
Certainty	Low	No indication of monitoring
	High	Indication that transfers are monitored on a random basis
Celerity	Low	Indication that sanctions will be imposed within six months after the error was detected
	High	Indication that sanctions will be imposed immediately after the error was detected

4.2. Measurements

The in-task behavior will be measured by the total number of errors committed by participants. Although, compliance research most likely relies on scenario-based approaches by presenting a scenario and measuring the prospective behavior, we chose an experimental setting to shed light on the interplay between formal sanction mechanisms and different kinds of human errors. Table 3 shows exemplary errors that can occur.

Table 3
Exemplary Medication Errors

Type of Error	Name	Dosage	Frequency
Original	Misoprolol	5mg	1-1-1-0
Data is omitted	-	-	-
Data is transferred inaccurately	Misoprolol	10mg	
Data is added additionally	Metformin	500mg	1-1-1-0

The sanction severity, certainty, and celerity constructs will be measured on a 7-point Likert-scale. Therefore, we will use previously validated items from the information security and compliance literature [19,21]. We plan to add the three control variables age, job experience, and resistance to change.

As the independent variables are reflective constructs, we will use the partial least square (PLS-SEM) method for analysis. In the first step, we will perform an assessment of the measurement model by evaluating the constructs' reliability (composite reliability and items' factor loadings) as well as the convergent and discriminant validity. The medications errors will be evaluated by their respective factors' relevance and will be tested for multicollinearity [22]. In the second step, the structural equation model will be assessed by performing a variance-based PLS approach and using the bootstrapping method [22].

Research Continuation

This research-in-progress paper introduces our identified research problem on medication errors and summarizes our research approach to answer the question of how organizational compliance mechanisms affect different kinds of data transfer errors in non-integrated IT-landscapes. With the study, we aim to contribute to the literature on the DT by investigating the relationship between perceived sanction severity, certainty, and celerity and the occurrence of various types of non-compliance. Furthermore, we aim to show how formal sanction mechanisms

can be used to prevent medical errors. Practitioners shall be able to use the results to define policies that help to prevent errors in manual data transfers. To validate our research model and research approach, we invite other researchers to provide feedback on our study.

References

- [1] J. S. Ash, M. Berg, E. Coiera, Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-related Errors, *Journal of the American Medical Informatics Association* 11 (2004) 104–122.
- [2] A. Alassaad, U. Gillespie, M. Bertilsson, H. Melhus, M. Hammarlund-Udenaes, Prescription and transcription errors in multidose-dispensed medications on discharge from hospital: an observational and interventional study: Errors in multidose-dispensed medications, *Journal of Evaluation in Clinical Practice* 19 (2013) 185–191.
- [3] World Health Organization, Medication Safety in Polypharmacy, World Health Organization, Geneva, 2019.
- [4] World Health Organization, Medication Without Harm – Global Patient Safety on Medication Safety, World Health Organization, Geneva, 2017.
- [5] E. Manias, S. Kusljic, A. Wu, Interventions to reduce medication errors in adult medical and surgical settings: a systematic review, *Therapeutic Advances in Drug Safety* 11 (2020).
- [6] R. Aron, S. Dutta, R. Janakiraman, P. A. Pathak, The Impact of Automation of Systems on Medical Errors: Evidence from Field Research, *Information Systems Research* 22 (2011) 429–446.
- [7] T. Heart, A. Zucker, Y. Parmet, J. Pliskin, N. Pliskin, Investigating Physicians’ Compliance with Drug Prescription Notifications, *Journal of the Association for Information Systems* 12 (2011) 235–254.
- [8] J. Callen, J. McIntosh, J. Li, Accuracy of medication documentation in hospital discharge summaries: A retrospective analysis of medication transcription errors in manual and electronic discharge summaries, *International Journal of Medical Informatics* 79 (2010) 58–64.
- [9] S. Trang, A. B. Brendel, A Meta-Analysis of Deterrence Theory in Information Security Policy Compliance Research, *Information Systems Frontiers* 21 (2019) 1265–1284.
- [10] R. Kohli, S. S.-L. Tan, ELECTRONIC HEALTH RECORDS: HOW CAN IS RESEARCHERS CONTRIBUTE TO TRANSFORMING HEALTHCARE?, *MIS Quarterly* 40 (2016) 553–573.
- [11] gematik, 2023. URL: <https://www.gematik.de/anwendungen/e-patientenakte>.
- [12] Barmer, 2023. URL: <https://www.barmer.de/gesundheitsverstehen/medizin/medikamente/wofuer-medikationsplan-1056422>.
- [13] M. Baehr, Traditionelle Arzneimittelversorgung in der Klinik, in: M. Maelzer, S. Melzer (Eds.), *Closed Loop Medication Management*, 1st. Ed. Medizinisch Wissenschaftliche Verlagsgesellschaft, Berlin, 2018, pp. 3–13.
- [14] Borycki E., Trends in Health Information Technology Safety: From Technology-Induced Errors to Current Approaches for Ensuring Technology Safety, *Health Informatics Research* 19 (2013) 69–78.
- [15] M. Kumar, G. Leroy, Factors Affecting Compliance with Alerts in the Context of Healthcare-related Emergencies, *Transaction on Replication Research* 3 (2021).
- [16] J. D’Arcy, T. Herath, A review and analysis of deterrence theory in the IS security literature: making sense of the disparate findings, *European Journal of Information Systems* 20 (2011) 235–254.
- [17] R. Paternoster, S. Simpson, Sanction Threats and Appeals to Morality: Testing a Rational Choice Model of Corporate Crime, *Law & Society Review* 30 (1996) 549–584.
- [18] J. P. Gibbs, *Crime, punishment, and deterrence*, 1st. ed., Elsevier, New York, 1975.
- [19] A. C. Johnston, M. Warketin, M. Siponen, AN ENHANCED FEAR APPEAL RHETORICAL FRAMEWORK: LEVERAGING THREATS TO THE HUMAN ASSET THROUGH SANCTIONING RHETORIC, *MIS Quarterly* 39 (2015) 113–134.

- [20] Y. Chen, K. Ramamurthy, K.-W. Wen, Organizations' Information Security Policy Compliance: Stick or Carrot Approach?, *Journal of Management Information Systems* 29 (2012) 157–188.
- [21] T. Herath, H. R. Rao, Protection motivation and deterrence: a framework for security policy compliance in organisations, *European Journal of the Association for Information Systems* 12 (2009) 106–125.
- [22] S. Trang, I. Nastjuk, Examining the role of stress and information security policy design in information security compliance behaviour: An experimental study of in-task behavior *Computers & Security* 104 (2021).

The Structure of Data Privacy Compliance

Alexandra Klymenko, Stephen Meisenbacher and Florian Matthes

Technical University of Munich, School of Computation, Information and Technology, Department of Computer Science, Boltzmannstr. 3, Garching, 85748, Germany

Abstract

Achieving Data Privacy Compliance involves a dynamic process requiring the expertise of many roles, particularly legal and technical experts, and it ultimately revolves around the goal of data protection, particularly in technical systems. While this goal may be clear, the inner workings and overall structure of the compliance process remain under-researched. In particular, the roles involved in the process of data privacy compliance and the nature of the interactions between them have not yet been investigated or formalized in a structured manner. In this work, we present such a structure, based on a series of interviews conducted with privacy professionals with varying responsibilities in compliance programs.

Keywords

Data privacy, privacy compliance, organizational structure

1. Introduction

With the growth in scrutiny placed upon data processing entities, particularly in light of recent regulations such as the General Data Protection Regulation (GDPR), the importance of proper privacy compliance programs has concurrently risen. Essentially, the demonstration of compliance with regulations involves the safeguarding of personal information via organizational and technical measures. In order for such measures to be successfully implemented, a series of (inter)actions and decisions must be carried out. Therefore, compliance is not an isolated action, but rather a *process*, in which multiple roles and responsibilities are involved.

In the changing landscape of data protection in response to rapidly advancing technologies and the regulatory response thereto, the process of privacy compliance has been continuously evolving. As such, little work has been performed to achieve a better understanding of such processes from an organizational perspective. This includes the different roles involved, their general categorization, the interactions between these roles, as well as the nature of such interactions.

This work presents the results of our initial investigation into structuring the process of privacy compliance with a focus on the implementation of *technical* measures [1]. In particular, we introduce the primary roles involved, whose responsibilities are crucial to the success of compliance programs, or the processes put into place to achieve compliance. These insights are obtained from a series of interviews with privacy professionals working in these processes. Next, we discuss the interactions taking place between these roles. Finally, as an additional new contribution to this extended abstract, we visualize our findings in a compliance structure, which we pose to be a representation of the general makeup of privacy compliance programs.

The structure of our work is as follows. Section 2 introduces the foundations of our work, upon which we built. Section 3 outlines our research design, the results of which are presented in

CIISR 2023: 3rd International Workshop on Current Information Security and Compliance Issues in Information Systems Research, co-located with the 18th International Conference on Wirtschaftsinformatik (WI 2023), September 18, 2023, Paderborn, Germany

✉ alexandra.klymenko@tum.de (A. Klymenko); stephen.meisenbacher@tum.de (S. Meisenbacher); matthes@tum.de (F. Matthes)

ORCID 0000-0001-7485-2933 (A. Klymenko); 0000-0001-9230-5001 (S. Meisenbacher); 0000-0002-6667-5452 (F. Matthes)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

Section 4, which culminates in our Privacy Compliance Structure (PCS). We conclude our work in Section 5, describing points of future work.

2. Background and Related Work

Modern privacy regulations, such as the GDPR or CCPA, establish guidelines for the responsible handling of personal data and require strict compliance, i.e., "ensuring adherence of an organization, process or (software) product to laws, guidelines, specifications and regulations" [2]. More specifically, the process of privacy compliance involves implementing various technical and organizational measures to ensure the protection of personal data, and, as such, it requires the expertise and involvement of specialists of various professional backgrounds. Research in the field of regulatory compliance of software systems, such as by Maxwell et al. [3], has demonstrated that software engineers cannot independently reason about compliance requirements. Likewise, Altman et al. [4] promote a hybrid legal-technical approach to privacy protection, arguing that without legal input, the technical solutions developers create to achieve regulatory compliance may prove ineffective in delivering strong privacy protection and risk non-compliance with regulations. Usman et al. [5] also point out the need to coordinate and align different compliance activities and roles, highlighting some of the difficulties that occur in the process. In this regard, Klymenko et al. [6] describe eight concrete challenges in the technical-legal interactions that occur in the process of data privacy compliance. In the following extended abstract, we aim to structure the roles and interactions involved in this process, in order to support further research on addressing current issues and advancing the overall efficacy of privacy compliance.

3. Methodology

We perform qualitative research following Grounded Theory methodology as described by Hoda et al. [7] by conducting semi-structured interviews with privacy experts in legal and technical sectors. To initiate the study, we developed an interview guide consisting of a pre-defined set of questions that were sent to participants in advance. These questions were designed to gain insights into the various roles, responsibilities, and interactions involved in the implementation of privacy requirements. We recorded and transcribed each interview, subsequently analyzing the data through coding and constant comparison, guided by thematic analysis [8]. This process continued until sufficient information was collected, allowing us to conclude the study. In particular, the main stopping criterion was the observation of a saturation of themes in our thematic analysis. In total, 9 legal experts and 7 technical experts were interviewed. Table 1 provides further information on the interviewees, where participant ID suffixed with a 'T' denotes a technical expert, 'L' a legal, and 'LT' a technical/legal expert. *Exp.* denotes years of experience (parentheses indicate experience specifically in privacy), and *Dur.* the duration of the interview in minutes. The interviews lasted for approximately an hour and were conducted via Zoom.

4. The Privacy Compliance Structure

4.1. Roles

Through the interview study, three overarching categories of roles were highlighted, each of which participates in the privacy compliance process from a different angle. In the interview discussion, the goal was not only to learn about the role of the interviewee, but also about relevant interactions with other roles, including the responsibilities of these further roles. Following the interviews, these roles and their responsibilities were extracted, and they are outlined below.

Table 1
Interview study participants

ID	Position	Organization	Exp.	Dur.
I1-T	Privacy Engineer	Large US media conglomerate	10+ (1)	54
I2-2	Privacy/Security Architect	Large German multinational software corporation	6	52
I3-L	Privacy and cybersecurity lawyer	US law firm	20+	32
I4-T	Privacy Engineer	Large US multinational tech company	5+ (4)	70
I5-LT	DPO, Managing Director	Small German data protection software company	4	50
I6-T	Software Architect	Large German multinational tech conglomerate	3	50
I7-L	Lawyer/external DPO	Small German data privacy company	20+	55
I8-L	Group Data Protection Counsel	International financial technology corporation	6	60
I9-T	Privacy Engineer	Large US Tech Corporation	8	65
I10-LT	Legal Counsel	Global Web Consortium	25	60
I11-L	Legal Counsel	German-based digital privacy consulting firm	3	50
I12-L	DPO	German-based consulting firm	20 (3)	55
I13-T	Security and Privacy Architect	Large German multinational tech conglomerate	3	60
I14-L	Compliance Officer	British-based news corporation	3	50
I15-T	Privacy Engineer	Chinese multinational tech corporation	15	60
I16-L	Legal Associate	Indian-based law firm	3	55

4.1.1. Legal

The first major role category consists of *legal experts*, which generally refer to practicing lawyers or legal associates, often specialized in data privacy or cybersecurity. These roles can be filled internally, or contracted to external legal counsel. The responsibility of these roles is to provide legal support and advice regarding the legal requirements set forth by relevant laws and regulations for privacy compliance.

Another type of legal role comes with the *consultant*. In the absence of or in supplement to lawyers, consultants are important to providing expert knowledge of the proper handling of data, also in light of relevant legal requirements. While consultants often may be a key point of the compliance process, these roles are often not practicing lawyers, showing that legal expertise may come in multiple forms.

A third type of legal role, although not explicitly legal, is that of the *compliance team*. Headed by a *Compliance Officer*, the compliance team is responsible for spearheading compliance activities, including assessing operational risk and ensuring that compliance steps are properly in line with data privacy principles. While the presence of such teams can be observed in practice, it is not clear how widespread such a unit is.

4.1.2. Technical

The process of privacy compliance also requires the involvement of technical experts, especially for the translation of legal requirements into technical solutions. As such, various technical roles were revealed to be involved, which we categorize under the term of *Development*. The first of these is the *product team*. Led by the *Product Owner*, the team is responsible for identifying and assessing potential privacy risks in a proposed product or project. Particularly with the Product Owner, this role becomes the "first point of contact" for ensuring privacy in a specific system.

Also important to the technical side of privacy compliance is the general role of *architects*. *Software Architects* are responsible for the design, rather than implementation, of systems, and it is in this phase where privacy matters must be first addressed. Architects may sometimes be specialized as *Privacy (and Security) Architects*, and also *Enterprise Architects* in larger organizations. In these roles, the design of privacy-preserving systems is crucial in the larger context of privacy compliance.

The role of experts in the *Development* category has been studied in the literature [9][10], speaking to the importance of such roles on the implementation level of privacy requirements. However, both works also make note of the challenges regarding both motivation and ability of

technical experts to comply with privacy regulations. Concrete challenges include tensions with legal [9] or a perceived lack of support [10], suggesting starting points for future work.

Further abstracted from the implementation of technical systems is the role of *management*, which is separated from development. Such roles are tasked with the leadership and direction with regards to compliance programs, ultimately giving the green light for compliance programs. A concrete role often mentioned was the *Chief Information Security Officer (CISO)*. The literature also points to the newer role of *Chief Privacy Officer (CPO)* [11]. While such roles do not participate in the implementation of technical measures for privacy compliance, they are indirectly involved via their interaction with Architects and members of the product team.

4.1.3. Go-Betweens

As a final group existing between the legal and technical roles, we define the category of *Go-Betweens* consisting of roles of a more hybrid nature. Concretely, we identified two important roles falling under this categorization.

Particularly since the GDPR came into effect, the role of *Data Protection Officer (DPO)* has become central to the compliance process. The DPO is appointed to steer and monitor all privacy compliance activities, such as Data Protection Impact Assessments (DPIA) or awareness-raising programs. DPOs can be internally filled, or also served by external persons such as consultants. At the core of the responsibilities of this person lies the task of liaising between the letter of the law and how this is interpreted in practice for organization-specific data processing activities. With this, it is clear that the DPO serves a hybrid role, with a leaning towards the legal aspects. Nevertheless, the multi-faceted responsibilities of the DPO include having technical knowledge, as noted by Ciclosi and Massacci [12].

The relatively novel, yet rapidly developing role of *Privacy Engineer* is tasked with bridging the legal-technical gap by providing technical expertise. While the discipline of Privacy Engineering is not yet fully mature, the general role involves creating trust by protecting privacy in technical systems and shaping organizational policy to facilitate this. Privacy engineers do not necessarily need to be experts in the specifics of a system; it is their expertise in privacy design principles that provides value to privacy compliance. As noted by Gürses and del Alamo [13], the rise of this role comes with an increasing need to bridge privacy research and practice, particularly in aligning privacy goals with legal policy, thus solidifying the role of a *Go-Between*.

4.1.4. Other Roles

The interviews also revealed other "external" supporting roles. On the regulatory side, supervisory authorities were indicated as important stakeholders. Within an organization, Marketing, HR, and Sales were also mentioned. Finally, the customer was often listed as a stakeholder, as the "true benefactor" of compliance efforts.

4.2. Interactions

Between the three main categories of roles in privacy compliance come many interactions, consisting of the *Technical-Technical*, *Technical-Legal*, and *Legal-Legal* nature. In defining these types of interactions, we utilize our categorization of roles as introduced above. Thus, an interaction is reported if an interviewee mentioned a relationship between roles within the legal or technical role categories, or between the two.

4.2.1. Technical-Technical

Within the technical sphere of an organization, interactions occurring in the process of privacy compliance come in two general forms: vertical and horizontal. Vertical interactions refer to the role of management positions, which pass down decisions regarding data privacy, as well as

provide support in liaison with legal experts. Architects may often interact with managers for matters regarding compliance.

More horizontally, privacy engineers will often interact with other technical roles, such as software engineers, in order to provide guidance and policy for the implementation of compliant systems. It should be noted that interactions also occur within a team, e.g., when privacy engineers of various specializations consult with one another.

4.2.2. Technical-Legal

As the process of privacy compliance is inherently interdisciplinary, technical-legal interactions are commonplace and almost necessary to ensure the success of compliance programs. The interviews highlighted that many of these interactions take place between legal experts such as lawyers and the technical leadership of an organization, i.e., management. In these interactions, the interpretation of legal requirements becomes very important for leaders to make informed decisions regarding privacy.

For other technical roles, not in management positions, the main point of contact regarding legal matters is the DPO. As access to the DPO is much more readily available than to lawyers, the DPO becomes a *de facto* Go-Between in bridging the technical-legal divide. This type of interaction, initiated by technical roles, can be useful to engineers with legal questions, or, specifically, to validate that legal requirements are being fulfilled in technical implementations.

An interesting question arises whether technical-legal interactions occur in the other direction, where legal roles initiate contact. While this type of interaction was not reported, it can be best observed in the close work of DPOs with Privacy Engineers or in cross-functional teams. Nevertheless, a more in-depth exploration of the legal roles in privacy compliance presents an interesting opportunity for future work.

A final type of technical-legal interaction comes with the existence of *cross-functional* teams. While this is not a common practice, such teams consist of members from different departments, including those that are more technically or legally oriented. These teams serve as an ideal place for cross-disciplinary exchange.

4.2.3. Legal-Legal

While legal-legal interactions were reported, such as those between consultants and lawyers, they are not expounded upon, as this is not the focus of our work.

4.3. The Structure

Based on the insights obtained regarding the roles, responsibilities, and interactions of the privacy compliance process, we propose a structure to illustrate the process and the dynamics within. As such, we have created the Privacy Compliance Structure (PCS), presented in Figure 1.

The PCS was created as follows. Firstly, the descriptions of the roles held by the interviewees were consolidated, serving as the foundation for the structure. Next, the interviews were analyzed for mentions of other roles in privacy compliance; more importantly, the nature of interaction between this newly mentioned role and the role of the interviewee was noted. This helped to form the basis of the interactions seen in Figure 1. Finally, the major sections in the structure, e.g., *Development*, were positioned to facilitate the nature of the interactions, as described above. For example, the vertical interaction between Management and Development is clear, and the close work between Privacy Engineers and DPOs is also reflected.

In Figure 1, solid single-directional arrows represent designated reporting lines, whereas single-directional dashed lines represent indirect reporting lines (where direct interaction is rare). Solid bi-directional arrows denote exchange rather than reporting lines, and hollow bi-directional lines indicate that two roles may be served by the same person. Finally, a cyclical arrow denotes when exchange occurs within a team.

5. Conclusion

In this work, we outline the components that are part of the process of privacy compliance, namely the roles, their responsibilities, and the interactions within. Using these findings, we construct a Privacy Compliance Structure that mirrors the above-mentioned components. In presenting this structure, we hope to provide structure to the dynamic process of privacy compliance, particularly in the implementation of technical measures.

As suggestions for future work, we see that a validation of the proposed structure is necessary to boost the generalizability of the model. To accomplish this, compliance programs of various organizations, from large and small, as well as those in differing domains, should be studied. Targeted case studies may be a useful approach for this. In addition, as the field continues to evolve, the inclusion of novel and currently not considered roles should be emphasized. As such, we hope that our base structure can be refined, updated, and evaluated.

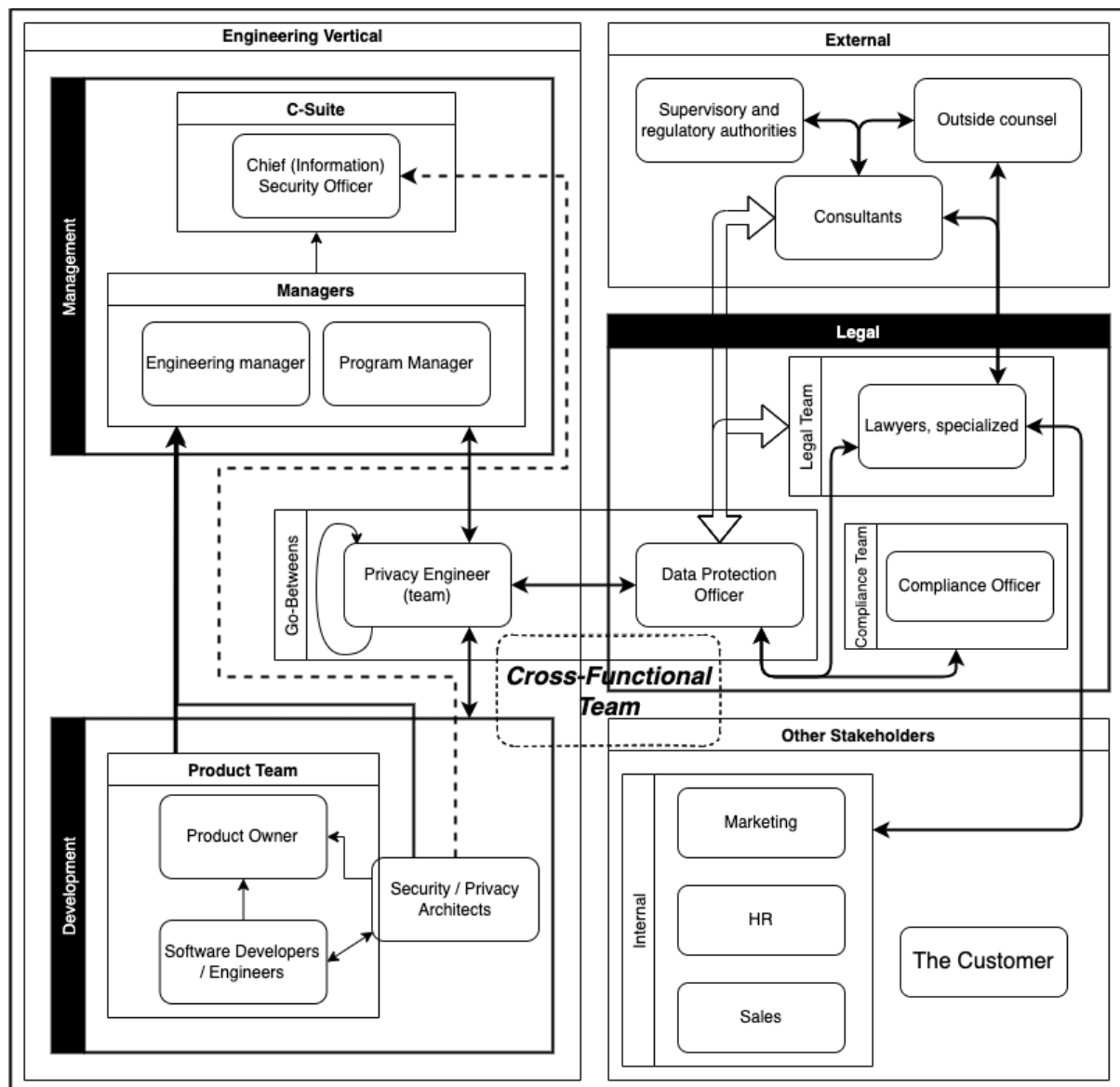


Figure 1: The Privacy Compliance Structure

Acknowledgements

This work has been supported by the German Federal Ministry of Education and Research (BMBF) Software Campus grant LACE 01IS17049.

References

- [1] O. Klymenko, O. Kosenkov, S. Meisenbacher, P. Elahidoost, D. Mendez, and F. Matthes, Understanding the Implementation of Technical Measures in the Process of Data Privacy Compliance: A Qualitative Study, in Proceedings of the 16th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement, 2022, pp. 261–271.
- [2] O. Akhigbe, D. Amyot, and G. Richards, A systematic literature mapping of goal and non-goal modelling methods for legal and regulatory compliance, Requirements Engineering, vol. 24, pp. 459–481, 2019.
- [3] J. C. Maxwell, A. I. Antòn, and J. B. Earp, An empirical investigation of software engineers' ability to classify legal cross-references, in 2013 21st IEEE International Requirements Engineering Conference (RE), 2013, pp. 24–31.
- [4] M. Altman, A. Cohen, K. Nissim, and A. Wood, What a Hybrid Legal-Technical Analysis Teaches Us About Privacy Regulation: The Case of Singling Out, SSRN Electronic Journal, 2020.
- [5] M. Usman, M. Felderer, M. Unterkalmsteiner, E. Klotins, D. Mendez, and E. Alégroth, Compliance requirements in large-scale software development: An industrial case study, in International Conference on Product-Focused Software Process Improvement, 2020, pp. 385–401.
- [6] O. Klymenko, S. Meisenbacher, and F. Matthes, Identifying Practical Challenges in the Implementation of Technical Measures for Data Privacy Compliance, AMCIS 2023 Proceedings. 2, 2023.
- [7] R. Hoda, J. Noble, and S. Marshall, Grounded Theory for Geeks, in Proceedings of the 18th Conference on Pattern Languages of Programs, Portland, Oregon, USA, 2011.
- [8] V. Braun and V. Clarke, Using thematic analysis in psychology, Qualitative research in psychology, vol. 3, no. 2, pp. 77–101, 2006.
- [9] K. Bednar, S. Spiekermann, and M. Langheinrich, Engineering Privacy by Design: Are engineers ready to live up to the challenge?, The Information Society, vol. 35, no. 3, pp. 122–142, 2019.
- [10] M. Tahaei, K. Vaniea, and A. Rashid, Embedding Privacy Into Design Through Software Developers: Challenges and Solutions, IEEE Security & Privacy, vol. 21, no. 1, pp. 49–57, 2023.
- [11] M. Shawosh, M. Bantan, and F. Belanger, Chief Privacy Officer role and organizational transformation in the digital economy, in ICIS 2022 Proceedings, 2022.
- [12] F. Ciclosi and F. Massacci, The Data Protection Officer: A Ubiquitous Role That No One Really Knows, IEEE Security & Privacy, vol. 21, no. 1, pp. 66–77, 2023.
- [13] S. Gürses and J. M. del Alamo, Privacy Engineering: Shaping an Emerging Field of Research and Practice, IEEE Security & Privacy, vol. 14, no. 2, pp. 40–46, 2016.