Burmeister, Fabian; Kurtz, Christian; Schirmer, Ingrid

Article — Published Version

Governing information privacy in data ecosystems with architectural thinking

Electronic Markets

Provided in Cooperation with:
Springer Nature

Suggested Citation: Burmeister, Fabian; Kurtz, Christian; Schirmer, Ingrid (2025) : Governing information privacy in data ecosystems with architectural thinking, Electronic Markets, ISSN 1422-8890, Springer, Berlin, Heidelberg, Vol. 35, Iss. 1, https://doi.org/10.1007/s12525-025-00808-5

This Version is available at:
https://hdl.handle.net/10419/323898

Electronic Markets (2025) 35:67
https://doi.org/10.1007/s12525-025-00808-5

RESEARCH PAPER

Governing information privacy in data ecosystems with architectural thinking

Fabian Burmeister[1] · **Christian Kurtz**[1] · **Ingrid Schirmer**[1]

Received: 22 March 2024 / Accepted: 5 June 2025

Abstract

This study addresses the challenge of governing privacy within data ecosystems by integrating architectural thinking (AT) into the discourse. Organizations are increasingly embarking on complex data-sharing initiatives, often encompassing personal data, raising heightened privacy concerns and regulatory obligations. With transparency obscured by complexity, there is a pressing need for systematic approaches to decompose data ecosystems. Leveraging AT, traditionally applied in intra-organizational contexts, this article proposes extending its application to the ecosystem level. By triangulating qualitative data from case studies and expert interviews, key privacy concerns of both business and regulatory stakeholders in data ecosystems are unveiled. Based on these concerns, the study develops and demonstrates a comprehensive data ecosystem architecture meta-model as a foundation for AT to govern privacy from both a business and regulatory perspective. The contributions bridge existing gaps in understanding and addressing privacy concerns within data ecosystems, offering stakeholders a systematic approach for analysis and mitigation.

Keywords Architecture · Ecosystem · GDPR · Meta-model · Privacy · Regulation

JEL Classification M15

Introduction

With the growing demand to leverage data for economic value, businesses are increasingly engaging in data ecosystems to acquire, share, process, and consume data (Oliveira et al., 2019). In these ecosystems, businesses trade data as a commodity to co-create value, whether through collaborative provision of data-driven services or via platforms that facilitate data sharing (Attard et al., 2016; Azkan et al., 2020). Data ecosystems have become a central focus within information systems (IS) research, addressing the complexities of data-driven networks that involve socio-technical relations and interactions among diverse actors, technologies, and resources (Heinz et al., 2022; Kurtz & Burmeister, 2024; Möller et al., 2024). While data in these ecosystems may originate from sensors or machines, a significant portion comprises personal data generated by individuals through smart devices, social media, or mobile apps, raising serious privacy concerns (Gopal et al., 2018).

The surge in personal data availability through data ecosystems and the allure of monetization, particularly through personalized advertising, raise privacy challenges for business and regulatory stakeholders. On the one hand, stricter privacy regulations (e.g., California Consumer Privacy Act (CCPA), General Data Protection Regulation (GDPR)) force business stakeholders (e.g., CEOs, app developers) to balance the benefits of personal data exploitation with privacy risks more than ever before (Li et al., 2019; Martin et al., 2019). Additionally, the dependence on collaboration within increasingly complex data ecosystems challenges businesses in maintaining transparency about data processing activities and in tracking data flows to customers, partners, and

Responsible Editor: Cinzia Cappiello

✉ Fabian Burmeister
fabian.burmeister@uni-hamburg.de

Christian Kurtz
christian.kurtz@uni-hamburg.de

Ingrid Schirmer
ingrid.schirmer@uni-hamburg.de

[1] Department of Informatics, Universität Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany

Published online: 17 July 2025

individuals (Crain, 2018; Tene & Polonetsky, 2013). On the other hand, the rise of widespread privacy breaches (e.g., Facebook and Cambridge Analytica), along with continuous advances in information technology (IT) to obtain personal data, requires regulatory stakeholders (e.g., data protection officers (DPOs), legislators, privacy lawyers) to gain more detailed insights into data ecosystems, assess privacy violations more efficiently, and align privacy regulations (Conger et al., 2013; Kurtz & Burmeister, 2024). Consequently, scholars call for novel approaches to unravel the manifold socio-technical relations that constitute data ecosystems, thereby fostering the anticipation, prevention, and analysis of privacy violations (Crain, 2018; Nissenbaum, 2019; Oliveira et al., 2019).

Following this call, this study draws on the paradigm of architectural thinking (AT) (Winter, 2014) to help both business and regulatory stakeholders cope with the growing complexity of data ecosystems. We argue that the intertwinement of organizations in data ecosystems, comprising interlaced privacy statements, opaque data flows between actors, and distributed IT landscapes processing personal data, calls for AT that encourages these stakeholders to "think and act architecturally" (Winter, 2014, p. 362) to manage privacy concerns in increasingly nontransparent data ecosystems. So far, AT has been discussed in the intra-organizational context, where it is defined as "the way of thinking and acting throughout an organization that considers holistic, long-term system aspects as well as fundamental system design and evolution principles in everyday decision making, which is not restricted to architects" (Aier et al., 2015, p. 390). Since AT should enable stakeholders to decompose and analyze socio-technical relations, a key for leveraging AT is to build up modeling competences (Sandkuhl et al., 2018; Winter, 2014). Such modeling competences require a holistic meta-model that enables a common language among stakeholders and identifies "what to model" by covering major elements and relations of a phenomenon (Frank, 2014). Moreover, following Knackstedt et al.'s (2014) research agenda for conceptual modeling in law, we argue that, in the context of privacy, a twofold business and regulatory perspective on data ecosystems is necessary, covering tasks related to both the application and creation of law. While business stakeholders focus on applying the law (e.g., weighing decisions about data sharing or third-party service integration against potential privacy violations), regulatory stakeholders must continuously create or revise laws based on insights gained from data ecosystems and privacy breaches. Thus, to extend the scope of AT to data ecosystems and the privacy context alike, this study aims to develop a data ecosystem architecture meta-model that helps business and regulatory stakeholders handle privacy concerns. Our research questions are as follows:

RQ1: *What are key concerns of business and regulatory stakeholders about privacy in data ecosystems?*
RQ2: *Which elements and relations should be integrated in a data ecosystem architecture meta-model to address the identified concerns and thereby leverage architectural thinking in the privacy context?*

Recently, modeling ecosystems were highlighted as a viable lens of analysis for the privacy field (Elrick, 2021; Kurtz et al., 2018). However, existing ecosystem meta-models (Belo & Alves, 2021; Burmeister et al., 2019a, 2019b; Oliveira et al., 2018) do not focus on privacy concerns and the underlying socio-technical relations for personal data sharing. There is a great lack of research clarifying how scholars and practitioners can apply the ecosystem perspective to cope with increasing privacy concerns. We contribute to these research gaps by extending the scope of AT through a data ecosystem architecture meta-model and showing how AT can reveal privacy issues. Following Lagerström et al.'s (2009) stakeholder-oriented meta-modeling approach, we start with identifying key concerns (Niemi, 2007) of business and regulatory stakeholders. For this purpose, we triangulate qualitative data gained from a multiple case study (Yin, 2009) and 14 expert interviews (Myers & Newman, 2007). Based on these concerns, we develop a coherent meta-model. The remainder of this paper is structured as follows. In the next section, we summarize related research. Then, we explain our methodology. Subsequently, we present the privacy concerns identified and demonstrate our meta-model. Finally, we discuss our results and draw a conclusion.

Related research

Data ecosystems

A data ecosystem is defined as "a loose set of interacting actors that directly or indirectly consume, produce, or provide data and other related resources, including software, services, and infrastructure" (Oliveira et al., 2019, p. 604). Refining this notion, Gelhaar et al. (2021) emphasize that data ecosystems encompass the collective generation, processing, and utilization of data across diverse actors, all aimed at co-creating value. Möller et al. (2024) underline that data ecosystems emerge around various data intermediaries, ranging from more open ones accessible to anyone (e.g., data marketplaces) to more restricted ones limited to users who meet specific governance policies (e.g., data trusts) and to those focused on societal or public issues (e.g., data collaboratives). Moreover, data ecosystems emerge around data spaces representing "decentralized data infrastructures designed to enable data-sharing

scenarios across organizational boundaries by implementing mechanisms for secure and trustworthy data sharing" (Möller et al., 2024, p. 6). Additionally, Curry and Sheth (2018) suggest distinguishing between four data ecosystem types based on the level of control over key data and participant interdependence: organizational (centrally controlled with independent participants), distributed (centrally controlled but requiring participant collaboration), federated (decentralized and voluntary collaboration for a predefined goal), and virtual data ecosystems (decentralized, with coalitions forming to pool resources). This study's focus is on identifying key elements constituting data ecosystems from a privacy perspective, according to interviewed business and regulatory stakeholders. The intention is not to be restricted to specific data intermediaries or data ecosystem types, as we seek to leverage AT for the analysis and mitigation of privacy concerns across various configurations, while keeping it lightweight and not limited to IT experts (Aier et al., 2015; Winter, 2014).

The research stream on data ecosystems is increasingly growing. Recent research created a morphology (Azkan et al., 2020) and a taxonomy (Gelhaar et al., 2021) that allow characterizing data ecosystems in different settings. In another work, Azkan et al. (2022) demonstrate the usefulness of modeling for handling data ecosystems. Specifically, they use the e3-value modeling language to capture different ecosystem roles, such as data providers, app store providers, and ecosystem orchestrators, along with their value co-creation. Additionally, Scheider et al. (2023) provide a reference system architecture for human-centric data ecosystems with related design principles. However, much literature regarding data platforms (Otto & Jarke, 2019), data marketplaces (Koutroumpis et al., 2017), and data ecosystems (Aaen et al., 2022; Oliveira et al., 2019) stresses the unexpected use and recombination of data in different contexts (Jarvenpaa & Essén, 2023). Several "challenges arise, including conflicting views about what constitutes appropriate behavior in relation to the exchange of data, given societal and individuals' privacy and security concerns" (Jarvenpaa & Essén, 2023, p. 3). In this regard, Aaen et al. (2022) distinguish between the dark and bright sides of data ecosystems, where function, stakeholder, and data creeps lead to adverse consequences for individuals. In line with this, Lis et al. (2023) highlight the increase in regulatory demands as a major challenge and call for inter-organizational data governance.

Above all, Oliveira et al. (2019) highlight several research gaps. In particular, research on modeling data ecosystems, especially their socio-technical architecture, remains underdeveloped. Despite their potential for value co-creation, data ecosystems pose heightened privacy risks due to increased accessibility of personal data. There is a notable lack of research addressing these privacy concerns. Tene and Polonetsky (2013) and Elrick (2021) underscore this gap, emphasizing the need for further research on privacy in data ecosystems.

Personal data and privacy

Several data protection and privacy regulations exist (e.g., GDPR, CCPA). In this study, we rely on the definitions and concepts of the GDPR as an exemplary regulation that spans multiple countries and significantly impacts not only the economy and society but also other regulations (Li et al., 2019). According to the GDPR (2016), personal data are "any information relating to an identified or identifiable natural person ('data subject')" (Art. 4 (1)). Personal data appear in various types (e.g., contact, health, location data) and are processed for manifold purposes, such as profiling users, analyzing customer behavior, enabling personalized service experiences, or complying with legal requirements (Acquisti et al., 2016; Purtova, 2018). Organizations monetize personal data through direct sales to third parties, such as advertisers or data brokers (Gopal et al., 2018; Myers West, 2019). Conversely, organizations acquire personal data to deepen their customer knowledge or even de-anonymize individuals (Crain, 2018). In this regard, privacy can be defined as "the ability of individuals to control the terms under which their personal information is acquired and used" (Culnan & Bies, 2003, p. 326). Informed consent, as the predominant legal basis that organizations rely on when processing personal data (GDPR, 2016, Art. 6), implies that individuals have transparency about data processing, including third-party recipients and their purposes, when deciding what personal data can be collected, shared, analyzed, or otherwise processed. However, the assumption of creating transparency via privacy statements is doubtful, as information about third parties is often hidden in unclear statements (Cate & Mayer-Schönberger, 2013; Gopal et al., 2018). Only a few privacy models address this issue by generalizing what actors may gain access to personal data (Benson et al., 2015; Conger et al., 2013). Moreover, there is a research gap on specifying the socio-technical relations between these actors that ultimately lead to privacy violations (Ananny & Crawford, 2018; Kurtz et al., 2018). Privacy violations result in adverse consequences for individuals, including social, physical, psychological, and prosecution-, career-, resource-, and freedom-related consequences (Karwatzki et al., 2017).

Architectural thinking and meta-model extensions

Architecture can be defined as "the fundamental concepts or properties of a system in its environment embodied in its elements, relationships, and in the principles of its design and evolution" (ISO 2011, p. 2). To capture the enterprise architecture (EA), where the "system" represents an organization, research and industry frameworks propose EA meta-models

to structure EA elements across logical layers. For example, the meta-model of the EA framework TOGAF defines four layers with corresponding elements: business (e.g., goals, processes), data (e.g., data entities), application (e.g., IS services), and technology (e.g., physical components) (The Open Group 2018). A meta-model is "the description of the set of notations and concepts used to define a model" (ISO 2002, p. 4) and "specifies the modeling elements used within another (or the same) modeling notation" (IEEE, 2003, p. 3). In contrast, a model is "an external and explicit representation of a part of reality as seen by the people who wish to use that model to understand, change, manage, and control that part of reality" (Pidd, 2009, p. 12). In other words, according to the Object Management Group (2002), a meta-model "defines the language for expressing a model" (p. 10), while a model is an instance of a meta-model. The relationship between meta-models and models is extensively discussed by, e.g., Da Silva (2015) and Kühne (2006). In the EA context, EA meta-models support the derivation of models representing current (as-is) or planned (to-be) states of an EA. These models address concrete stakeholder concerns, which are "interests related to the development of EA, its use, and any other aspects that are important to one or more stakeholders" (Niemi, 2007, p. 2). Based on EA models, the EA management (EAM) performs tasks related to strategic planning and business-IT alignment (Saat et al., 2010).

However, researchers state that EAM cannot develop its full potential, as its influence is often limited to IT departments (Aier et al., 2015; Winter, 2014). Thus, EAM should evolve to AT as a less formalized and utility-centered approach that supports non-IT experts to analyze, plan, and transform fundamental structures (Winter, 2014). An overview of differences between EAM and AT is provided by Winter (2014). Research on AT is scarce and has focused on the intra-organizational level. For example, Aier et al. (2015) examine determinants, challenges, and adoption mechanisms of AT in the financial sector. Horlach et al. (2020) propose design principles for AT to support organizational agility. In prior work, we leveraged AT for inter-organizational tasks in large-scale e-government projects (Burmeister et al., 2019b), aiming to foster a collaborative mindset among stakeholders (i.e., not just IT experts) to understand how different parts of an ecosystem (e.g., actors, IS, processes) are interrelated. AT must be adapted in the privacy context to facilitate the decomposition and analysis of data processing and sharing within data ecosystems, thereby uncovering the causes of privacy violations.

Due to our focus on meta-models, previous work on extending their scope through a privacy or data ecosystem lens is notable. For example, Demchenko et al. (2014) modeled key components along the big data lifecycle, including analytics tools, cloud services, and access control. Erraissi and Belangour (2018) propose meta-models for data source and ingestion layers in Hadoop, covering data types (e.g., GPS), data formats (e.g., unstructured), and analytics steps (e.g., compression). To support GDPR compliance, one article proposes a privacy-driven EA meta-model with data processing and security layers next to business and IT layers (Burmeister et al., 2019c). However, modeling socio-technical relations and data sharing is limited by the intra-organizational perspective of EA. Hence, another study proposes a meta-model for digital transformations in business ecosystems, covering domain (e.g., influences), actor (e.g., actor classes), collaboration (e.g., services), and IT (e.g., interfaces) layers (Burmeister et al., 2019a). Oliveira et al. (2018) provide first steps toward a data ecosystem meta-model by suggesting actors, roles, relationships, and resources as elements. However, their research does not focus on privacy and personal data sharing with third parties. Key enablers for the latter, according to the boundary resources concept, are application programming interfaces (API) and software development kits (SDK) (Eaton et al., 2015).

In summary, research on data ecosystems, privacy, and AT is still in its infancy. No previous work embeds AT within the inter-organizational privacy context, as we do through a data ecosystem architecture meta-model designed to help business and regulatory stakeholders manage complex privacy concerns. While prior research addresses privacy from the EA perspective (Burmeister et al., 2019c), existing ecosystem meta-models lack a detailed focus on modeling privacy-related, socio-technical relations between actors.

Methodology

In our study, we followed Lagerström et al.'s (2009) stakeholder-oriented meta-modeling approach, which draws on stakeholder concerns to create architectural fragments and integrate these into a coherent meta-model. To identify diverse privacy concerns and increase the validity of our results, we triangulated data gained from an explorative multiple case study (Benbasat et al., 1987; Yin, 2009) and expert interviews (Myers & Newman, 2007). We demonstrated and evaluated our results through practical application and focus groups (see Fig. 1).

Data collection

In a long-term study, we collected privacy scandals (hereinafter termed cases) in a case study database to explore GDPR deficits regarding irresponsible data handling in ecosystems. For this article, we analyzed only four of these cases (see Table 1), as the intention was to get an initial idea of privacy-critical socio-technical relations in data ecosystems and stimulate discussion during subsequent interviews. To select representative cases, we relied on theoretical

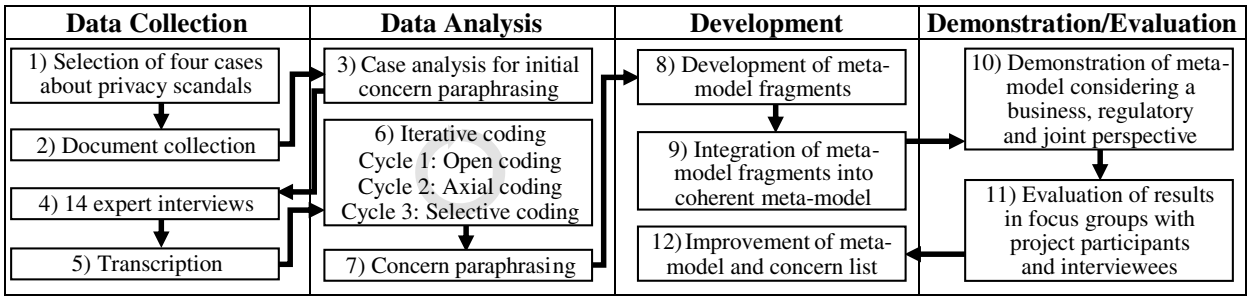

Fig. 1 Methodology

Table 1 Overview of cases

No	Title	Description
1	Facebook and Cambridge Analytica	The developer of the Facebook app This Is Your Digital Life (TIYDL) indicated to collect profile data for research purposes but sold the data to Cambridge Analytica that processed the data to target over 87 million Facebook users with political ads (documents collected: 35, exemplary news article: cnet.co/3k5c7ci)
2	Red Shell and gaming software	Device fingerprints collected by the spyware Red Shell, which is integrated in dozens of games such as Civilization VI and The Elder Scrolls Online, are used to measure the effectiveness of ads displayed in the web browsers of players (documents collected: 21, exemplary news article: bit.ly/355nhcM)
3	Healthcare app Ada	The provider of the healthcare app Ada stated that health data would not be shared with third parties, but a network traffic analysis revealed that data on symptoms and devices had been transmitted to companies like Amplitude or Facebook (documents collected: 12, exemplary news article: bit.ly/2zREn0y)
4	Waitlist app OpenTable	The waitlist app OpenTable, used by more than 47,000 restaurants, shares personal data like contact details, location data, and user preferences with other Priceline-owned sister companies, such as Kayak and Booking.com, and with advertisers (documents collected: 13, exemplary news article: cnet.co/3eAXr3b)

replication logic, wherein cases are predicted to yield contradictory results (Benbasat et al., 1987; Yin, 2009). Theoretical replication enabled us to consider cases with heterogeneous characteristics (e.g., diverse actor roles, data types, and adverse consequences) and increase our study's generalizability. As such, we did not limit the case study to specific data ecosystem types, such as those described by Curry and Sheth (2018). In detail, we selected the cases for multiple reasons. First, they involve data ecosystems heavily reliant on personal data but vary across industries (e.g., healthcare, gaming), types of privacy violations (e.g., purpose misuse, spyware), and actors (e.g., Facebook as platform provider, restaurants as service providers). Second, they represent large-scale ecosystems with thousands of users and various third parties, exhibiting manifold socio-technical relations. Third, they were widely covered in the media, providing rich research material.

For each case, the data collected involved meticulous searches across prominent news platforms (i.e., The New York Times, Wired UK, BBC, CNet, and ZDNet). We aggregated content from these platforms and expanded our data through backward searches using linked articles to uncover additional content from other platforms. In total, we collected 81 documents, including 71 news articles, two official responses from actors, five blog entries, and three technical reports. Analyzing this data provided valuable insights into the socio-technical elements and relations in data ecosystems, enabling us to anticipate associated privacy concerns (see section "Data Analysis").

However, to mitigate potential bias in subjective interpretations, we complemented the case study with 14 expert interviews (see Table 2), following Myers and Newman's (2007) seven guidelines. For example, to ensure our interviews "represent various 'voices'" (Myers & Newman, 2007, p. 17), our selection of interviewees covered both business and regulatory stakeholders, spanning heterogeneous roles (e.g., advisory, managerial, technical) and workplaces (e.g., public, commercial). Additionally, we considered that all interviewees are engaged in daily tasks related to personal data in organizations embedded in data ecosystems, such as data-driven businesses and privacy law firms. Due to our exemplary focus on the GDPR, all

Table 2 Overview of interviews

No	Interviewee role	Industry	Main topics of interest/longest discussed	Length/words
B1	Data analyst	Advertising	User profiling, personalization of ads, data acquisition	73 min/6039
B2	CEO	Software	Distribution via app stores, data-driven business models	57 min/5382
B3	Enterprise architect	Consumables	Personal data in EA models, GDPR compliance in EA	61 min/5481
B4	Data analyst	Finance	Personal data processing in lending and online banking	34 min/3092
B5	App developer	Software	Implementation of SDKs in apps, programming of APIs	72 min/5756
B6	CIO	Public service	Problems in ensuring IT security, data exchange formats	54 min/4782
B7	Enterprise architect	Railway	Modeling of interfaces, collaboration with departments	43 min/3523
B8	App developer	Mobile games	In-app purchases and ads, restrictions through GDPR	39 min/3184
R1	Privacy lawyer	Public service	Purpose limitation, general deficits of informed consent	69 min/5531
R2	DPO	Healthcare	Storage of sensitive data, IT security measures	48 min/4727
R3	Privacy lawyer	Software	Purpose changes through third parties, GDPR loopholes	52 min/4823
R4	Privacy lawyer	Insurance	Processing of personal data to calculate insurance terms	43 min/4241
R5	DPO	Mobile games	Contract design, transparency lack in privacy statements	28 min/2755
R6	Legal adviser	Consulting	Barriers to privacy compliance, usefulness of modeling	35 min/3012

interviewees operate in the German-speaking region, with three-quarters having studied either business administration, computer science, IS, or law.

Considering Myers and Newman's (2007) guideline for flexible interviews, we relied on a semi-structured interview guide with open-ended questions (see Fig. 2), facilitating in-depth discussions while accommodating the interviewees' specific interests. We referred to the data and concerns from our case study to inspire discussions (e.g., to brainstorm privacy issues using the cases). Prior to the interviews, we shared the interview guide and a summary of our research objectives to ensure a common understanding. All interviews were conducted, recorded, and transcribed by the same researcher. Subsequently, this and another researcher coded the data using MAXQDA. Following Myers and Newman's (2007) confidentiality guideline, the material was anonymized beforehand and shared only between these two researchers.

Data analysis

The data analysis aimed to identify major privacy concerns of business and regulatory stakeholders and gain a comprehensive understanding of socio-technical elements and relations constituting data ecosystems. The two researchers triangulated the data from the case study and transcriptions through a qualitative content analysis (Mayring, 2014). Starting with the case study, the researchers separately reviewed the data collected for each case and inductively coded socio-technical elements, such as actors (e.g., "data broker"), personal data (e.g., "location data"), and technical components (e.g., "iPhone). Relations were open-coded as well (e.g., "shared with third parties," "contractual agreement"). Additionally, a priori codes from the literature (e.g., adverse consequences, boundary resources) deductively guided the analysis. With minor coding differences, the researchers discussed, refined, and paraphrased the codes to obtain preliminary concerns (Lagerström et al., 2009; Niemi, 2007). For this, we were inspired by the structure for

Fig. 2 Excerpt from interview guide

Area of interest	Exemplary questions
Interviewee's role within data ecosystem	What are your primary tasks, particularly those related to personal data?
	How does your role influence data flows within your organization's ecosystem?
Challenges and solutions concerning privacy compliance	What key challenges do you face in ensuring privacy compliance within your organization's data ecosystem or for clients?
	How does your role influence data flows within your organization's ecosystem?
Information requirements pertaining to personal-data processing	What do you consider to be key elements and relations within data ecosystems?
	What types of information do you require to effectively manage personal data?
	To what extent do you desire more transparency about personal-data flows?
Prevailing practices in modeling data ecosystems	To what extent do you model your organization's data ecosystem?
	What modeling languages or tools do you use, and are there limitations?
Role of architectural thinking	In your opinion, will data ecosystems become even more complex in the future?
	How necessary is architectural thinking and decomposing data ecosystems?

formulating concern-related questions according to Lagerström et al. (2009). In detail, we used our codes as keywords for elements, attributes, and relations, and translated them into stakeholder concerns in the form of questions. Examples are given in Fig. 3. Although this initial analysis was subjective, the results gave an impression of privacy concerns about data ecosystems and inspired the interviews.

Following Saldaña's (2015) advice that multiple coding cycles are required for rigorous data analysis, the two researchers subsequently analyzed the transcriptions in three coding cycles. The first coding cycle was inspired by the case analysis, as induction and deduction were combined as well. A coding scheme (Mayring, 2014) was set up, which included not only codes from the literature (e.g., "API" and "SDK" as boundary resources) but also the codes identified in the case study. While many text passages could be coded via the scheme, the open coding revealed new codes, such as "purpose change" or "de-anonymization." In the second coding cycle, we used axial coding to group the codes into broader, theme-focused categories (Saldaña, 2015). This resulted in 14 categories, such as "collaboration" and "security measures." For example, the codes "contractual agreement" and "service co-creation" were assigned to the category "collaboration." In the third

Text quote* (from multiple case study and expert interviews)	**Code*** (coding cycle 1)	**Category*** (coding cycle 2)	**Concern area** (coding cycle 3)	**Derived concern*** (code paraphrasing)
"It is difficult to say where our ecosystem starts and ends. I would say, in the first place, you must know your direct environment, meaning your biggest **customers** and **partners**."	Actor	Actors	Actor composition	Which actors are interacting in a focal data ecosystem?
"If we sold the data to Google, we both would be the **controller** and they would still be the **processor** for us."	Role			Which roles do specific actors in the data ecosystem have?
"To be fully compliant with the GDPR, we have to treat all people as **data subjects**, with all their rights and our legal obligations, which involves quite a lot of effort."				
"I've lost track of **who supports us with which services**. I hope the departments know who they are cooperating with."	Service co-creation	Collaboration		Which services are co-created by which actors?
"The firm **combines data** from multiple sources, including online information and polling, to build 'profiles' of voters."	Method	Data analysis	Data processing	Which methods does an actor use to analyze personal data?
"… then we **store all the data** in our central data lake."				
"We mainly use the data to **improve our advertising processes**. For example, to send product recommendations."	Benefit			Which business processes profit from data processing activities?
"Transferred data included not only **technical information** about the smartphone and operating system but also the **symptoms** entered by the user after logging in."	Shared data type	Data sharing		With whom does an actor share which type of personal data?
"… it can share data, **like where you prefer to sit**, with other restaurants and with third-parties …"				
"In our customer app, people have **many options**. They can decide, for example, whether to **share their location or not**."	Privacy settings	Applications	IT landscape	Which privacy settings are included in provided applications?
"… if they book the train ride through our app, we receive the necessary information from their **smartphone**."	Tablet	Devices		Which devices are involved in a data processing activity?
"I tapped on the **iPad** screen and then noticed something that made me realize this restaurant was doing more with my name and phone number than just telling me a table was ready."				
"Our clients often need help with the record of processing activities for the GDPR. This is quite challenging, as especially large companies often lose the overview which data they are actually processing and for **what specific purpose**."	Purpose determination	Purpose limitation	Regulatory compliance	Which purpose does a selected data processing activity have?
"… gained access to the data of millions of Facebook users and then may have **misused it for political ads** …"	Purpose change			To what extent do third parties infringe original purposes?
"With all the data exchange between companies, it is **very difficult to retrace the original purpose**"				
"… all these companies have the possibility to merge data from different sources into comprehensive dossiers on users, because the data is generally **not sufficiently anonymised**."	Anonymization	Security measures		How does an actor ensure that personal data are anonymized?

***Only a selection is shown.**

Fig. 3 Excerpt from coding scheme

coding cycle, we applied selective coding (Flick, 2009) to integrate the categories into more meaningful themes. For example, the categories "data acquisition" and "data analysis" were labeled as "data processing." This resulted in four selective codes, which we call concern areas. After each coding cycle, inter-coder agreement tests were conducted to reach a consensus by comparing the codes and recoding the data. Similar to the case study, the researchers aggregated text quotes assigned to identical codes by formulating them as questions representing stakeholder concerns, following Lagerström et al. (2009). For example, the code "role" of the category "actors" was framed as "Which roles do specific actors in the data ecosystem have?" Fig. 3 outlines our coding scheme.

Development, demonstration, and evaluation

Guided by Lagerström et al.'s (2009) stakeholder-oriented meta-modeling approach, we created our meta-model in two steps. First, for each of the 14 categories, we created meta-model fragments by extracting elements from the stakeholder concerns (see Fig. 4). To keep the meta-model streamlined for lightweight AT, we aggregated elements with similar content per category. We translated content from the concerns into attributes if the content could be assigned to elements. For example, "aggregate" and "analyze" both represent actions performed by data processing activities. Additionally, we equipped attributes with frequently mentioned codes. For example, the attribute "type" of the element "data object" was refined by codes such as "financial" or "health." We also added relations to interconnect the elements based on the concerns. We refrained from using EA modeling languages like ArchiMate, as they are limited to predefined intra-organizational elements and lack the flexibility to cover ecosystem elements and legal aspects. Instead, we used the flexible class diagram and modeling tool draw.io.

Second, to obtain one coherent meta-model, we merged our meta-model fragments by relating them to each other and ensuring consistency. This involved comparing the fragments' elements, removing redundant elements by aggregating their attributes, and aligning or adding relations between the elements. Inspired by EA meta-models (e.g., TOGAF) and our concern areas, we added layers to our meta-model that structure elements with similar content and ensure modularity. While all interviewees mentioned the elements "actor," "data object," and "data processing activity," indicating they may be core to data ecosystems, we intentionally did not designate any elements as mandatory for model creation. Instead, we designed our meta-model so that all elements are flexibly selectable to address individual concerns (i.e., concern-driven selection), and the meta-model is modifiable to accommodate specific purposes or data ecosystem types. In this regard, we argue that such flexibility is essential to align with Stachowiak's (1973) general model theory, according to which derived models should fulfill three criteria: (1) mapping, a model represents a real-world phenomenon; (2) reduction, it simplifies the original, omitting some details; and (3) pragmatism, it is useful and serves a purpose, replacing the original for a specific task.

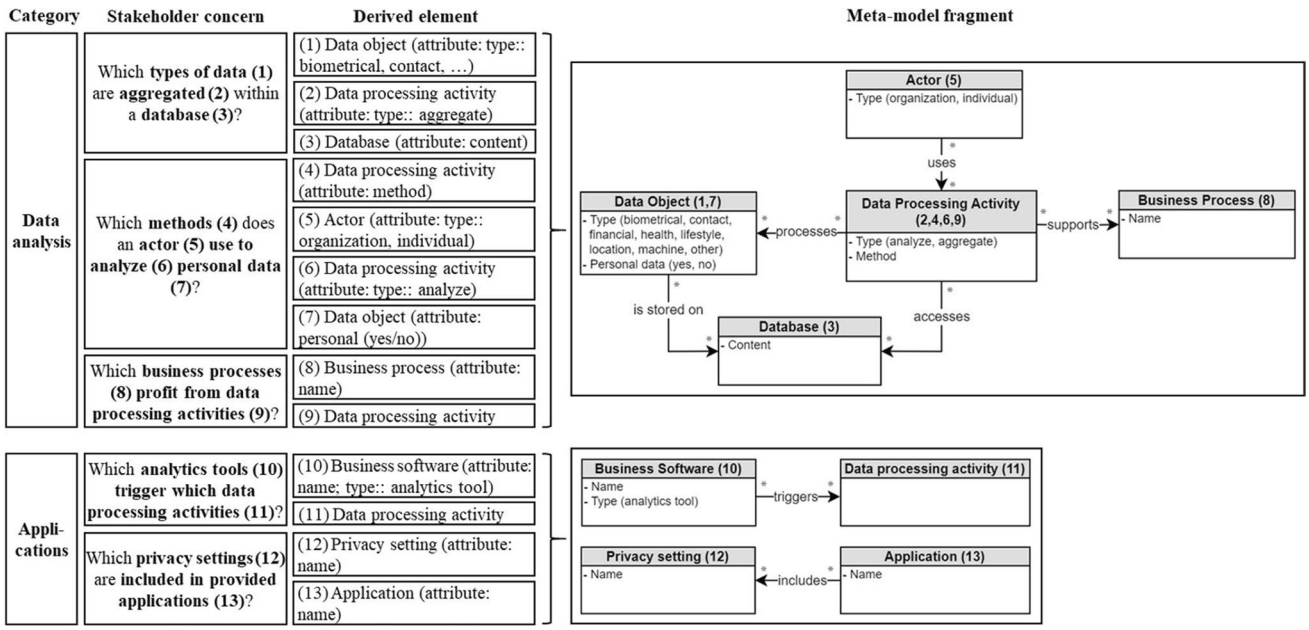

Fig. 4 Exemplary development of meta-model fragments

Figure 5 exemplifies how we merged the meta-model fragments from Fig. 4.

We demonstrated and evaluated the meta-model in a multiple case study (Yin, 2009), covering the perspectives of both business and regulatory stakeholders. First, from the business perspective, we supported privacy-related design decisions in the app development project NeighborBook (NB) by practically applying AT based on the meta-model. In multiple focus groups (Bélanger, 2012) with project participants, we modeled parts of the app's data ecosystem and evaluated the meta-model's usability, leading to several improvements. For example, we added the attribute "priority" to the meta-model's element "purpose" and a self-referencing relation to the element "application." Second, from a regulatory perspective, we collaborated with legal experts in a research project to model the data ecosystem of the Israeli contact tracing app HaMagen and reveal privacy issues. Based on their evaluation, we refined the meta-model by adding or renaming certain relations and attributes to improve clarity and usability. Third, integrating both business and regulatory perspectives, we modeled privacy scandals in the Reveal Mobile data ecosystem for an ex-post analysis. In a focus group with three previously interviewed experts (R2, R3, and R6), we evaluated the model and meta-model, leading to further refinements. For example, to simplify the meta-model, we removed elements such as "privacy settings" or "data type," as the experts found they were too fine-granular for an ecosystem perspective, limited the meta-model's usability, or overlapped with other elements and attributes. Overall, the experts appreciated the variety of concerns covered and the meta-model's layered structure. They found AT based on the meta-model useful to assess privacy in data ecosystems. Recognizing that no case selection is entirely free from bias, we followed theoretical replication logic (Benbasat et al., 1987; Yin, 2009) to address this issue. By selecting cases covering diverse domains (i.e., social networks, digital healthcare, and data monetization) and purposes (i.e., app development and legal assessment), we aimed to demonstrate the generalizability of our meta-model. Additionally, each case was chosen for its relevance, accessibility, and ability to provide empirical insights into privacy in data ecosystems. The first two cases were selected for the opportunity of in-depth stakeholder engagement through our project involvement, while the third reflected multiple privacy scandals covered in news articles, offering multiple real-world perspectives. While acknowledging potential sampling bias, our case selection balanced diversity and relevance, justifying our approach and findings in empirical realities and practical constraints, such as data access and time limitations.

Results

Our results leverage AT in the privacy context and extend its scope to data ecosystems. In the following, we first present the privacy concerns (RQ1) before demonstrating the application of AT based on our meta-model (RQ2).

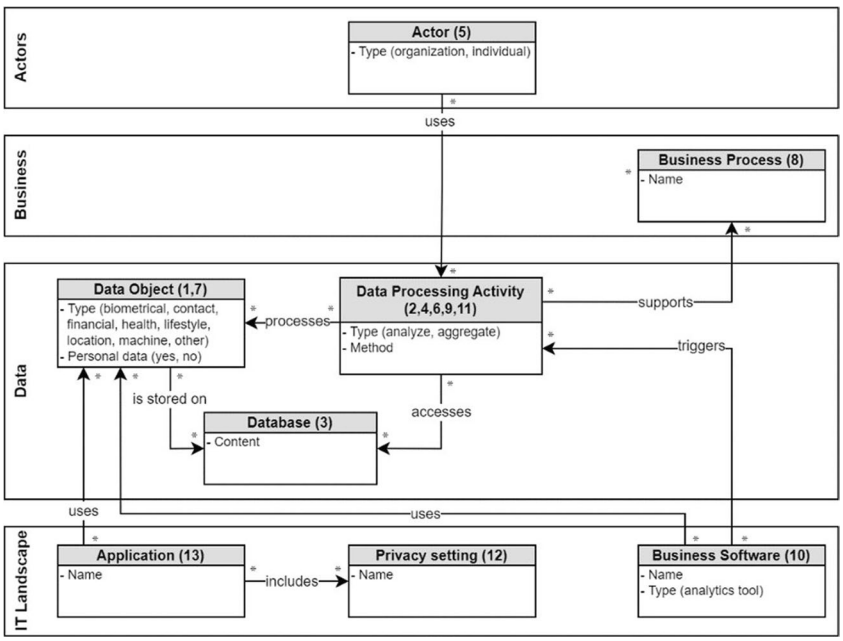

Fig. 5 Exemplary merging of meta-model fragments

Privacy concerns of business and regulatory stakeholders

The privacy concerns relate to (1) actor composition, (2) data processing, (3) IT landscape, and (4) regulatory compliance in data ecosystems, with each concern area comprising multiple categories from the axial coding.

Actor composition

Addressing privacy concerns from an architectural perspective can support several scenarios such as data-driven transformations, integrating external data sources into analytics, or assessing privacy violations. Stakeholders conducting AT must define the boundaries of the data ecosystem they are observing or altering. This can initially be done by identifying the relevant actors for a given scenario (C1), as noted by all interviewees (see Table 3).

Capturing the role of actors (C2) is essential to recognize their viewpoint and legal obligations. B2 said: "The GDPR divides companies into data controllers and data processors. This provides a basis for legal assessment. However, sometimes companies are both. For example, we use cookies from Google Analytics to measure our website traffic. They only process the data for us, so we are the data controller. If we sold the data to Google, we both would be the controller and they would still be the processor for us. It strongly depends on the situation." Estimating the extent to which an actor's business model relies on personal data (C3) supports understanding an actor's role in a data ecosystem. Our case study shows that privacy violations not only occur in the dyadic relationship between an individual and a service provider but often result from data sharing practices by a chain of actors. Therefore, understanding data ecosystems requires insights not only into intra-organizational data processing but also into multi-actor collaboration. This can be seen most clearly in contractual relationships (C5), which are influenced by the legal framework applicable in the countries where actors are incorporated (C4). For the regulatory stakeholders, access to contracts (GDPR, 2016, Art. 58 (1)) is important to obtain information about services (C6) and intended activities of individuals (C7). The interviewees also mentioned that capturing distribution channels is a good starting point to retrace privacy violations (C8) and emphasized the need to reconstruct how platforms are embedded in data ecosystems (C9), as these often trigger collaboration.

Data processing

Data processing is often no longer confined to individual organizations but occurs among collaborating actors, as analytics processes are increasingly outsourced and data have become a commodity. AT enhances transparency about data ecosystems by decomposing socio-technical relations. The interviewees raised multiple concerns as central for this task. Recognizing what types of personal data are processed and shared in a data ecosystem (C10) helps anticipate privacy violations and what information may be inferred (C11) (see Table 4).

To legally assess an actor's behavior, it is essential to know what personal data that actor collects (C12) and from which data sources (C13). A data protection officer stated: "In my opinion, modeling data types is essential to enable a comparison of data sharing networks. If a certain actor buys different data types from different actors, it may be possible to identify a black sheep that seeks to de-anonymize personal data" (R2). It is crucial to retrace with whom actors share personal data (C18), how often (C19), and if consent is obtained (C20).

Both business and regulatory stakeholders raised privacy concerns regarding the location of personal data storage (C14), the extent of data aggregation (C15), and the methods used to analyze personal data (C16). However, addressing these concerns requires in-depth knowledge of

Table 3 Privacy concerns about actor composition

Concern area	Category	No	Stakeholder concern	Source	
				B	R
Actor composition	Actors	C1	Which actors are interacting in a focal data ecosystem?	All	All
		C2	Which roles do specific actors in the data ecosystem have?	All	All
		C3	Whose business model is heavily reliant on personal data?	1 – 3,5,8	2,5
		C4	In which country does an actor have its legal seat?	2,5	All
	Collaboration	C5	How are actors contractually interconnected?	All	All
		C6	Which services are co-created by which actors?	1 – 3,5 – 7	1,2,4
		C7	To what extent do services deviate from intended user actions?		3,5
		C8	How do actors distribute applications and services?	2,4,5,8	1,4,5
		C9	To what extent are platforms involved in the data ecosystem?	1 – 5,8	1,3 – 6

Table 4 Privacy concerns about data processing

Concern area	Category	No	Stakeholder concern	Source	
				B	R
Data processing	Data types	C10	Which types of personal data are processed by which actors?	All	All
		C11	Which information can be inferred from selected data types?	1,4	2,4
	Data acquisition	C12	Which kind of personal data does a specific actor collect?	All	All
		C13	From what data sources does an actor acquire personal data?	All	All
		C14	Where does an actor store personal data?	All	All
	Data analysis	C15	Which types of data are aggregated within a database?	1,3 – 5	2,4,6
		C16	Which methods does an actor use to analyze personal data?	1,4,6	2 – 5
		C17	Which business processes profit from data processing activities?	1 – 3,6 – 8	3,6
	Data sharing	C18	With whom does an actor share which type of personal data?	All	All
		C19	How often does data sharing between two actors take place?	1,2,4,7	1 – 3,5
		C20	Has consent been given for a specific form of data sharing?	1,4,5,8	All

organizations that may not be easily accessible, as reflected by interviewees R2, R3, and R6 from legal practice. The interviewees also emphasized that business processes driven by data analytics should be captured (C17), clarifying purposes of data processing and supporting the record of processing activities (GDPR, 2016, Art. 30). The enterprise architect B3 stated: "We here in the EA department often help out when it comes to updating the record of processing activities. However, with all the data exchange with partners that the departments initiate on the fly, we increasingly lose track of which processes actually collect, share, or analyze personal data. We have to broaden our scope and consider partners and external data sources. The internal EA mindset is no longer sufficient today."

IT landscape

Decomposing data ecosystems requires knowledge of the IT landscape underlying data processing (see Table 5), which can initially be identified through the software that triggers data processing (C21) and is used by organizations (e.g., analytics tools) and individuals (e.g., mobile apps). Evaluating privacy settings (C22) and relating them to data flows help assess an actor's compliance with privacy by default (GDPR, 2016, Art. 32).

The respondents underlined that modeling of data ecosystems needs to reflect diversity of the devices involved in data processing activities (23), what personal data are transmitted by these devices and to whom in a specific case (C24), and what interfaces are involved in these data transactions (C25). A developer stated: "Many privacy violations result from a misuse of location data or a user's advertising ID transmitted via smartphones. These data are shared with advertisers or other third parties, allowing a form of tracking that is often not visible to users. So I would suggest that modeling and relating these devices to other elements is inevitable" (B5). Regarding this issue, the interviewees emphasized that many applications communicate via APIs (C26) and integrate predefined functions of SDKs (C27) provided by platforms such as Facebook. However, as demonstrated later (see Fig. 9), SDKs may access and share personal data in ways that bypass user consent.

Regulatory compliance

To assess privacy compliance in data ecosystems, it is necessary to identify the applicable regulations (C28) and

Table 5 Privacy concerns about IT landscape

Concern area	Category	No	Stakeholder concern	Source	
				B	R
IT landscape	Applications	C21	Which analytics tools trigger which data processing activities?	1,3,4,7	1,2,4,5
		C22	Which privacy settings are included in provided applications?	2,5,8	3 – 5
	Devices	C23	Which devices are involved in a data processing activity?	1 – 3,6	2,5
		C24	To whom are which data transmitted from individuals' devices?		4
	Technical interfaces	C25	Which technical interfaces does a data processing activity use?	4,3,6	
		C26	Which infrastructure components communicate via which API?	5	
		C27	Which SDKs are integrated in which applications?	5,8	

enforcement actors (C29) (see Table 6). For business stakeholders, this helps recognize legal obligations (e.g., the GDPR demands special guarantees for data transmissions outside the European Union (2016, Art. 44)).

According to all interviewees, capturing the purpose of data processing activities (C30) supports evaluating whether actors comply with purpose limitation (GDPR, 2016, Art. 5 (1)). By relating business processes, services, or data objects to data processing activities, deviant purposes can be revealed (C31). Furthermore, modeling how personal data are shared between actors and processed can illustrate the extent to which defined purposes are not met (C32). A privacy lawyer approved: "With all the data exchange between companies, it is very difficult to retrace the original purpose" (R4). The respondents also emphasized the need to model security measures regarding the anonymization (C33) and storage (C34) of personal data, which can reveal whether actors adequately ensure data protection and privacy by design (GDPR, 2016, Art. 32). Capturing the data an actor collects also helps to anticipate the risk of de-anonymization of personal data (C35). The interviewees added that comparing data processing activities (C36) and their purposes (C37) with those listed in a privacy statement can reveal gaps and non-compliance. This is especially critical when the modeling discloses third parties not listed in a privacy statement (C38). Besides regulatory compliance, ethical values (e.g., social welfare) must be considered in the assessment of privacy violations or actors' behaviors (C39). For example, under the GDPR (Art. 6), data processing can be lawful without consent to safeguard vital or public interests. In contrast, consent can result in adverse consequences that are not appropriate from an ethical point of view (C40).

Data ecosystem architecture meta-model to govern information privacy

The meta-model (see Fig. 6) consists of five horizontal and three vertical layers, incorporating socio-technical elements and relations derived from the stakeholder concerns and refined through evaluation. It serves as a basis for AT in privacy-related tasks and decision-making by facilitating the systematic decomposition and analysis of data ecosystems. Based on empirical data, the meta-model proposes key elements constituting data ecosystems from a privacy perspective while remaining lightweight, adaptable, and not limited to IT experts, in line with AT (Aier et al., 2015; Winter, 2014). Following a concern-driven approach that embraces reduction and pragmatism (Stachowiak, 1973), business and regulatory stakeholders can flexibly select elements from the meta-model to create models of real-world data ecosystems (i.e., instantiations of the meta-model) tailored to their specific privacy concerns (such as those outlined in Tables 3, 4, 5, 6) and purposes. The meta-model's adaptability allows stakeholders to modify, add, or remove modeling elements, attributes, and relations as needed.

As the topmost horizontal layer, the overarching data ecosystem layer includes elements such as regulations (e.g., GDPR) and ethical values that influence a data ecosystem's actors. An actor represents an organization (e.g., platform provider), which has an EA, or a group of individuals (e.g., app users), which has an individuals' architecture (IA). Actors interact through a boundary architecture (BA), described later in this section. The business layer contains business models and processes, collaboration-related elements such as services, and individuals' activities with adverse consequences resulting from a data ecosystem. Noteworthy is the importance of attributes. For example,

Table 6 Privacy concerns about regulatory compliance

Concern area	Category	No	Stakeholder concern	Source	
				B	R
Regulatory compliance	Legal framework	C28	Which privacy regulations apply to the data ecosystem?	2 – 4,8	1,2,4,6
		C29	Who is responsible for enforcing which regulation?	2	1,2,6
	Purpose limitation	C30	Which purpose does a selected data processing activity have?	All	All
		C31	Which data processing activities deviate from defined purposes?	1,3,4	All
		C32	To what extent do third parties infringe original purposes?	1	1,3
	Security measures	C33	How does an actor ensure that personal data are anonymized?	1,3 – 6	All
		C34	How does an actor ensure that personal data are stored securely?	2,3,5,6	2,3
		C35	Is a specific actor capable of de-anonymizing personal data?		2,3,6
	Privacy statements	C36	Is a data processing activity not defined in a privacy statement?		All
		C37	Are all purposes of data processing listed in a privacy statement?	1,4	All
		C38	Which third parties are not listed in a privacy statement?	1,4,5,7	1,3 – 5
	Ethical values	C39	To what extent do processing activities contradict ethical values?	1	3,4,6
		C40	What are the adverse consequences of a data ecosystem?	8	5,6

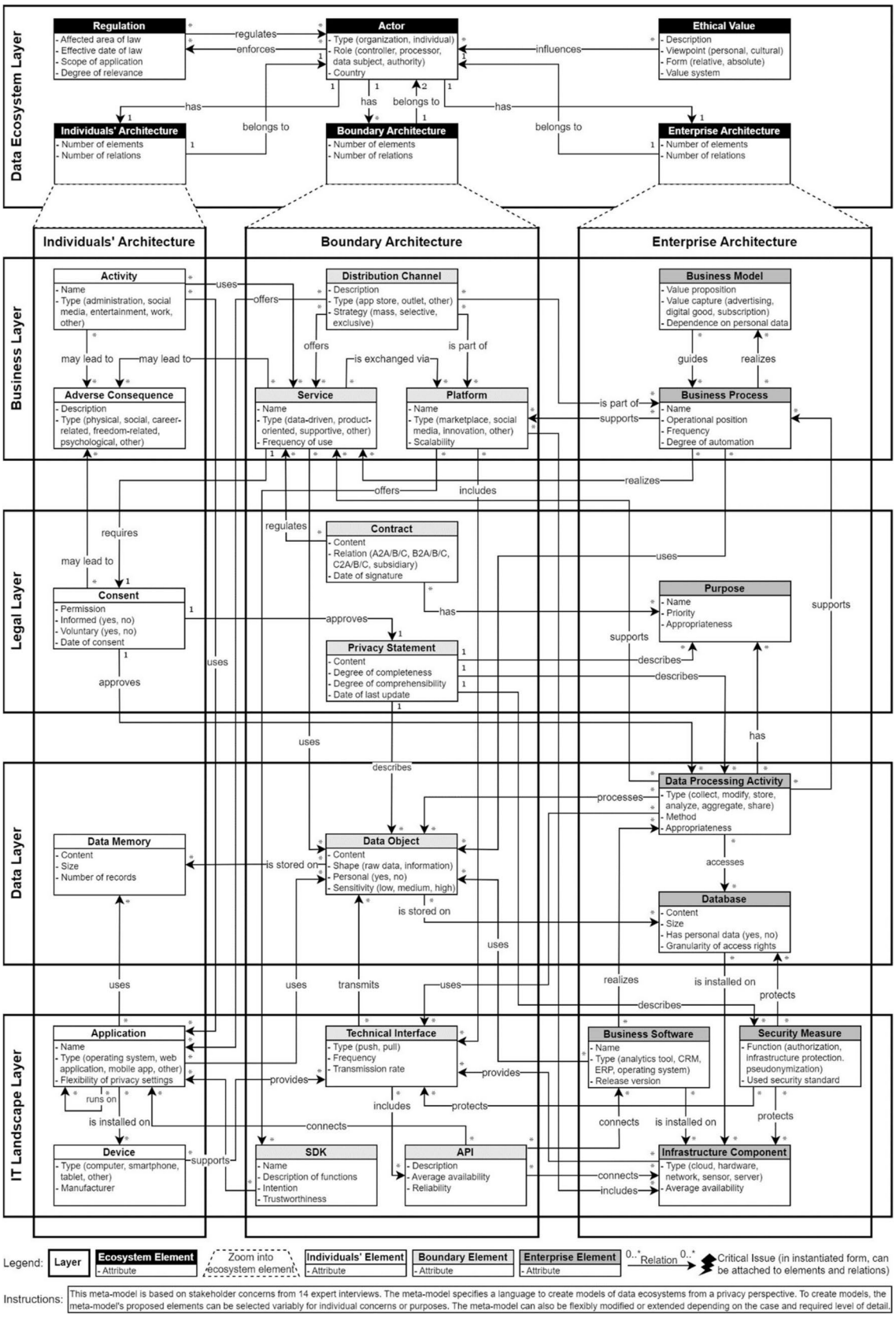

Fig. 6 Data ecosystem architecture meta-model to govern information privacy

the "type" of activities may help identify recurring patterns of data flows in data ecosystems (e.g., social-media-related activities are likely to involve platforms), while a business process's "degree of automation" indicates human involvement that influences privacy in different ways. While lower automation may introduce risks related to human intervention, such as unauthorized data access or resale of personal data (e.g., Facebook and Cambridge Analytica), higher automation, at least according to B2, B3, and R3, can also jeopardize privacy as highly automated processes tend to over-collect data and obscure data processing. Depending on the context, both extremes require careful privacy controls to mitigate risks. The legal layer includes consents, contracts, privacy statements, and data processing purposes for legally assessing data ecosystems. While the meta-model highlights consent and contract as legal bases by reflecting the interviewees' concerns (C5, C20), other bases, such as compliance with legal obligations or safeguarding vital and public interests (GDPR, Art. 6), can be added as needed. Additionally, consent and contract are legal bases in multiple regulations, such as the CCPA, enhancing the meta-model's generalizability. Next, the data layer represents data processing activities, data objects, and elements for data storage. For example, the "size" attribute of the database element may indicate large data lakes, while the "granularity of access rights" reflects the precision of data access control and associated exposure risks. The last layer covers the underlying IT landscape comprising software and hardware, technical interfaces, and security measures. Here, applications may represent individuals' mobile apps along with their "flexibility of privacy settings," indicating the extent to which users can control their data and manage privacy preferences. Such applications often integrate pre-defined functions from SDKs, with "trustworthiness" reflecting the reliability and security of these functions and their adherence to privacy standards (e.g., Apple requires SDK providers to comply with strict privacy policies). An example of an untrustworthy SDK is provided in the section "Demonstration" (see Fig. 9).

The vertical layers enable modeling of data ecosystems at high granularity by incorporating both intra- and inter-organizational perspectives. For the intra-organizational view, the meta-model zooms (indicated by dotted lines in Fig. 6) into both the EA and IA. Whereas the EA focuses on elements relevant to intra-organizational data processing, the IA summarizes involved elements of a group of individuals (e.g., users in a data ecosystem). For the inter-organizational perspective, the meta-model zooms into the interconnecting BA, which we define as the intersection of two organizations collaboratively processing or sharing personal data, of individuals and one organization co-creating data-driven services, and of individuals exchanging data with each other. In comparison to elements of the EA and IA, the elements of the BA are shared, i.e., jointly used by two actors and enabling collaboration. According to our meta-model, an actor has exactly one EA or IA but any number of BAs, i.e., one with each collaboration partner. We argue that the distinction between these architectures is necessary to comprehend data ecosystem complexity and consider today's interconnectedness and coalesced IS between organizations and individuals. Although these types of architectures can be used separately to address either intra- or inter-organizational concerns, most of the identified concerns show that AT in the privacy context requires a combination of both perspectives. To highlight elements and relations considered privacy-critical in a data ecosystem, a "critical issue" mark can be attached. In instantiations of the meta-model, when this mark is attached to elements or relations, it is represented by a flash symbol.

Demonstration of architectural thinking based on the meta-model

To demonstrate the usefulness of AT based on our meta-model for decomposing data ecosystems, identifying privacy-related critical issues, and mitigating privacy risks, we supported business and regulatory stakeholders in different tasks by creating and applying models (i.e., instantiations of the meta-model). By considering Moody's (2009) principles for visual representations during model creation, we aimed to enhance cognitive effectiveness within AT. Specifically, we leveraged the meta-model's layers (particularly the distinction between EA, IA, and BA) to support the principle of complexity management. In this regard, Moody (2009) states that "the most common way of reducing complexity of large systems is to divide them into smaller parts or subsystems" (p. 767). Furthermore, we argue that the flash symbol of the "critical issue" element, highlighted in the meta-model's legend, increases semantic transparency and visual expressiveness of our models by directing attention to key points from a privacy perspective. While we ensured that all semantic constructs of our meta-model are represented through a visual syntax (i.e., graphical symbols) according to Moody's (2009) principle of semiotic clarity, most elements in our models are depicted as rectangles due to strict adherence to the formal meta-model. While this supported the principle of graphic economy by minimizing the number of symbols, it also resulted in symbol overload. Symbol overload is a common issue in architecture modeling, as noted by Moody (2009) in the EA modeling language ArchiMate. To partially mitigate this issue in our models, we incorporated the layers, along with color-coded and labeled elements, guided by the principle of perceptual discriminability that seeks to differentiate symbols. Examples of how we step-by-step derived and refined models can be found in Burmeister et al. (2022).

In the following, our created models are presented. In line with the definitions provided in the "Related Research" section, these models represent parts of real-world data ecosystems, are instantiated from the meta-model, and serve specific purposes—in our case, the analysis and mitigation of privacy risks.

Business perspective

To demonstrate the usefulness of AT from the business perspective, we participated in the NB project in which we applied AT to support privacy-related design decisions. The aim of NB is to develop an online neighborhood social network. Users must provide their address during signup and receive a physical letter with a verification code to enter upon their first login. Correctness and precision of the address data are key to successful verification. Hence, the implementation of a third-party address autocompletion service was considered during development. Offered by an address autocompletion provider (AAP) and embedded in the NB web application via an external JavaScript tag, the service can provide real-time address completion based on characters entered in the address field. The service is offered for free, and its implementation requires no more than two lines of code. To assess privacy for the decision of implementing this service, the data ecosystem was modeled based on our meta-model. The resulting model revealed several issues (see 1a-3c in Fig. 7) related to the identified privacy concerns (see Tables 3, 4, 5, 6).

Issue 1a illustrates that the address autocomplete script transmits data from the NB web application to the AAP without involvement or control of the NB provider (C24, C25). As issue 1b shows, with this transmission, the AAP gains technical data, including the IP address, of the user's device and partial address data entered (C10, C12). Modeling the AAP's EA reveals issues pertaining to data processing activities. The identification of these issues is based on an analysis of the AAP's privacy policy, which reveals several BAs to third-party data buyers and sellers. Issue 2a shows that the AAP enriches technical and address data of users by purchasing further data (C13). Issue 2b clarifies that the AAP aggregates the data to create user profiles (C15, C30). Issues 2c and 2d show that the AAP analyzes the profiles to identify usage patterns (e.g., for ads or service improvement) and for other purposes (C30). Issue 2e shows that the AAP also shares the data with data buyers (C18). The concrete identities of third-party buyers and sellers remain unspecified (C38). Albeit these purposes can be extracted from the AAP's privacy policy, their relation to the core address autocompletion service is not immediately clear (C7, C31, C32). The model also identifies issues pertaining to additional effort for the service's implementation. Issue 3a shows the need for an agreement to be established between the AAP and NB (C5). As issue 3b indicates, NB must adapt its privacy statement to include the AAP as third-party data processor (C36-C38). Finally, following issue 3c, NB would need to gain consent from its users before the script is loaded on their device (C20).

AT and the model allowed for a holistic and well-informed design decision that considers potential privacy risks, implementation effort, and expected benefits for NB users. Consequently, the NB developers decided not to implement the autocompletion service. As such, AT helped prevent NB from becoming part of the AAP data ecosystem, thereby mitigating privacy risks by avoiding excessive personal data sharing. Similar design decisions, facilitated through instantiations of the meta-model, were made for an interactive mapping service, third-party email and physical mail services, and hosting providers, where AT also helped mitigate privacy risks.

Regulatory perspective

To demonstrate the usefulness of AT from the regulatory perspective, we provided the NGO Privacy Israel with

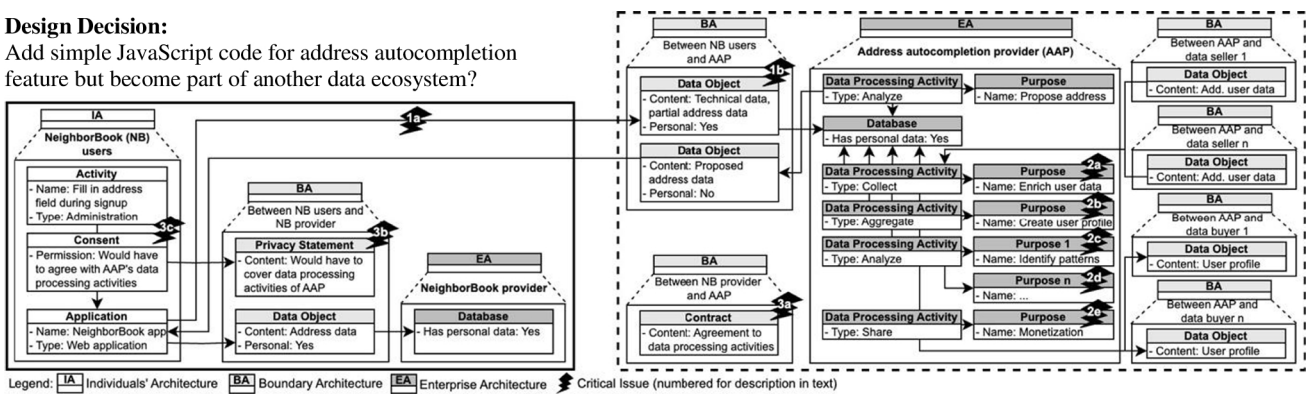

Fig. 7 Instantiation of the meta-model to support a design decision during app development

architectural knowledge about the COVID-19 contact tracing app HaMagen ("The Shield"), developed by the Israeli Ministry of Health (MoH). Its functionality was largely decentralized, with location data stored on users' mobile phones and checked against a central database of confirmed COVID-19 patients maintained by the MoH. The app's main function was to notify users if they had been exposed to a COVID-19-positive individual. No data was transferred to the MoH unless infected users consented to share it. While the Israeli public initially embraced HaMagen, with approximately 1,000,000 users 1 week after its launch, the MoH announced in November 2020 that it would cease further investment in the app.

To assess potential privacy impacts, we created a model of the app's data ecosystem based on its privacy policy, technical documentation, and scientific literature, revealing multiple critical issues (see 1–4 in Fig. 8). The model clarified that the MoH's epidemiological system, which is responsible for data processing of patients' sensitive data, remained a black box (issue 1). Moreover, based on the model, we became precise regarding knowledge gaps on data flows (C18), data sources (C13), data holders (C5, C10), and data controllers (C28, C29): What kinds of data can Systematics, a private provider of geographic IS, access (issue 2)? What specific data are inputs for and outputs of the epidemiological investigation unit (issue 3)? To what extent is the General Security Service involved in the HaMagen's data ecosystem (issue 4)? More details can be found in Burmeister et al. (2022).

AT not only helped identify these issues but also mitigate privacy risks when, together with Privacy Israel, we submitted an information request to the MoH. Although this request remains unanswered, we highlighted the diversity of privacy issues within the app's data ecosystem. With Privacy Israel's help, we informed stakeholders such as the MoH and app users about these issues and encouraged proactive measures to address them. As such, AT raises awareness of privacy among users and ensures that organizations like the MoH consider privacy risks more effectively in their decision-making, contributing to the mitigation of long-term harms.

Integrated business and regulatory perspectives

To demonstrate AT's value from both business and regulatory perspectives, we apply our meta-model to two additional privacy scandals from our case study database. The first case involves the gas station app GasBuddy (GB), which sold location data to Reveal Mobile, a geofencing marketing and analytics provider (exemplary link: cnet.co/33ai5CI), without users' knowledge. Reveal Mobile then shared this data with third parties, including data aggregators, insurers, and advertisers, leading to adverse consequences such as location-based ads. The second case concerns the weather app AccuWeather (AW), which also shared location data with Reveal Mobile. Even when users opted out, AW transmitted WiFi and Bluetooth data from user devices, allowing Reveal Mobile to approximate their locations (exemplary link: zd.net/2RfoPfI). We selected these cases as both involve Reveal Mobile, enabling us to capture them within a single meta-model instantiation and illustrate the complex data ecosystem of a data aggregator. The resulting model depicts several issues (see 1a–3 in Fig. 9).

Regarding the GB case, issue 1a shows that the privacy statement between GB and its users does not mention Reveal Mobile as a third party (C20, C36, C38). Shown by issues 1b and 1c, the GB provider shares user data for monetization

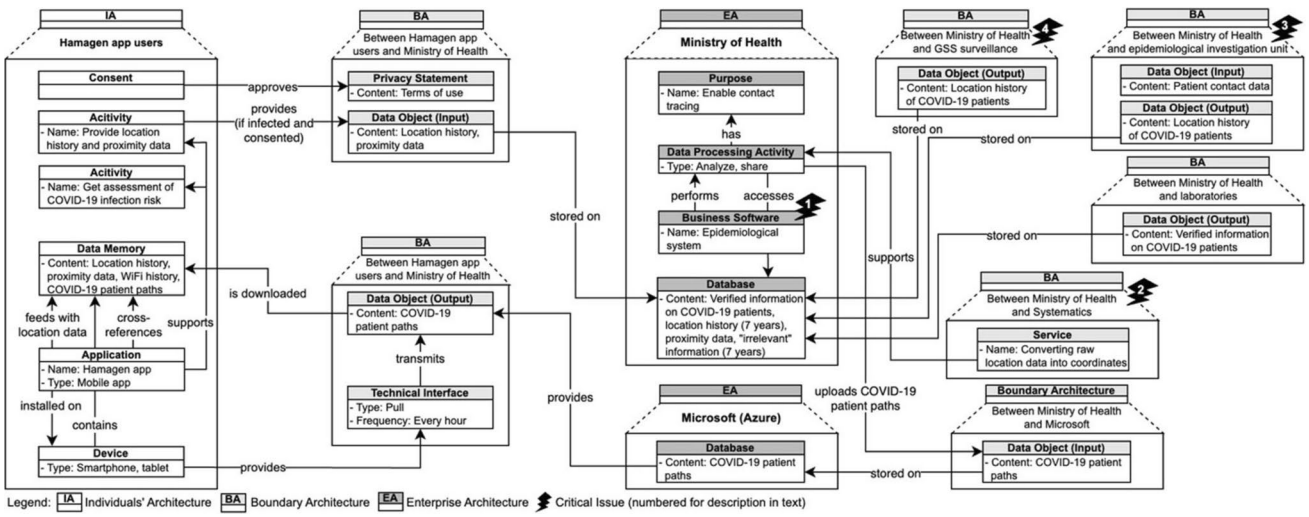

Fig. 8 Instantiation of the meta-model to assess privacy in a contact tracing app's data ecosystem

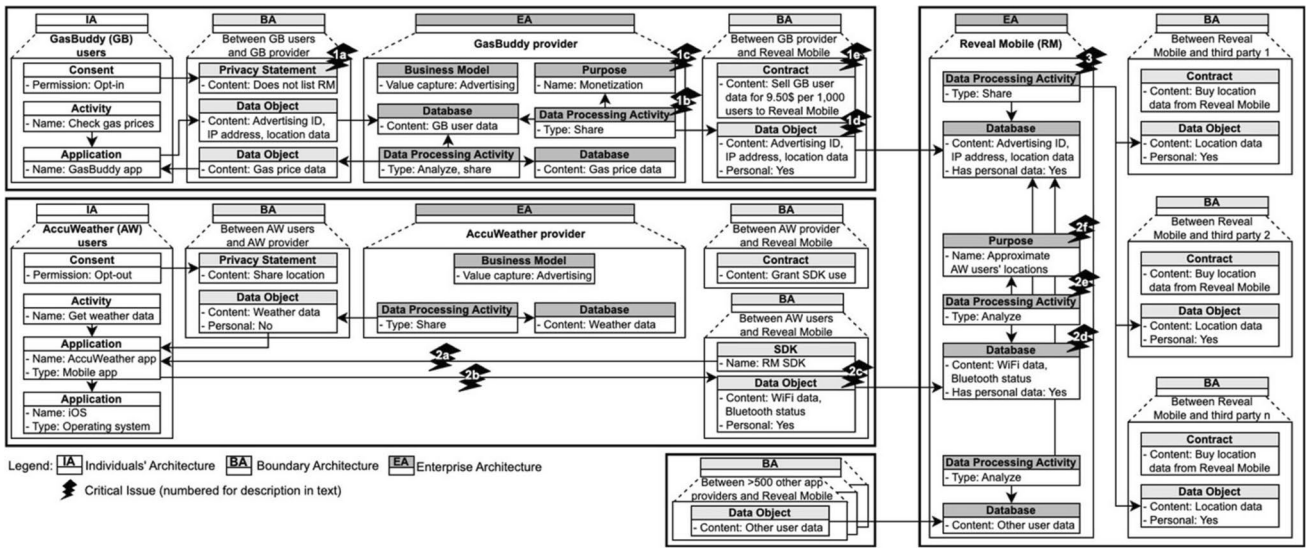

Fig. 9 Instantiation of the meta-model to assess privacy in a marketing and analytics provider's data ecosystem

purposes (C30, C37). Issue 1 d clarifies that the GB provider shares not only the location data with Reveal Mobile, but also the advertising ID and IP address of users' devices (C24). Issue 1e shows the contract between the GB provider and Reveal Mobile, about which users have not been informed (C5, C7). To highlight the critical issues in the AW case, we only modeled the opt-out alternative of users. Here, issue 2a shows that the AW app includes the Reveal Mobile SDK (C27). Issues 2b and 2c clarify that the AW app transmits users' WiFi data and Bluetooth status received from the iOS operating system to Reveal Mobile via the SDK (C18, C20). Issue 2 d illustrates that Reveal Mobile stores these data (C14). Based on the stored WiFi data and Bluetooth status, Reveal Mobile approximates AW users' locations (C11), shown by issues 2e and 2f. Finally, issue 3 points out that the location data, whether received directly as in the GB case or approximated based on technical data as in the AW case, is shared with third parties (C18). More than 500 apps use the Reveal Mobile SDK.

The model reveals important issues from both a business and regulatory perspective. First, it reveals how data aggregators like Reveal Mobile acquire personal data through poorly informed consent or SDKs. In the GB case, consent is questionable as Reveal Mobile was not mentioned in the privacy statement. Similarly, the AW case challenges informed consent, as user locations were inferred despite opting out. Legislators should address technical backdoors from third-party components like SDKs and clarify responsibilities. Second, the model underscores ecosystem complexity, where aggregators collect data from numerous apps and partners, blurring ecosystem boundaries. Third, while the GDPR (2016, Art. 26) governs dyadic relationships and joint-controllerships, transparency is needed to understand responsibilities among joint controllers.

To support regulatory action, we forwarded our findings on obscured third-party access and legal gaps—uncovered by decomposing Reveal Mobile's ecosystem through AT—to the European Commission for GDPR evaluations. In this regard, AT helped mitigate privacy risks by deriving policy recommendations for regulators, substantiating them with real-world examples, and encouraging authorities to identify and penalize such issues.

Discussion

Theoretical contributions

Our study makes several theoretical contributions. There is a lack of research on extending meta-models through a privacy-, data-, or ecosystem-oriented lens. Albeit some meta-models suggest data source (Erraissi & Belangour, 2018) or data processing layers (Burmeister et al., 2019c), they are limited by an intra-organizational perspective. In contrast, ecosystem meta-models do not focus on privacy concerns (Belo & Alves, 2021; Burmeister et al., 2019a, 2019b; Oliveira et al., 2018). Our meta-model contributes to this research gap by combining intra- and inter-organizational perspectives with business, legal, data, and IT landscape layers. While Oliveira et al.'s (2018) meta-model includes actors, relationships, roles, and resources as elements of data ecosystems, our meta-model encompasses multiple layers with corresponding elements, attributes, and relations. A notable feature of our meta-model is its

consideration of a legal layer. This aspect acknowledges the importance of legal contexts in shaping the behavior of actors within data ecosystems. Hereby, we follow Sandkuhl et al.'s (2018) call to "open up new domains for enterprise modeling – e.g., for conceptualizing modeling methods for the legal/compliance domain" (p. 77). By addressing privacy concerns of business and regulatory stakeholders, we contribute to the concern, stakeholder, and model scope dimensions proposed by Sandkuhl et al. (2018).

Furthermore, we demonstrate that AT based on our meta-model helps decompose the various socio-technical elements of data ecosystems, while modeling languages such as e3-value focus on roles and values (Azkan et al., 2022), and ArchiMate is restricted by its EA elements. We also argue that modeling to-be architectures of data ecosystems during IS design helps minimize different forms of creep with adverse consequences (Aaen et al., 2022; Karwatzki et al., 2017). As shown in our example of NB, architecture modeling can indicate stakeholder creep when potential third parties pose privacy risks. Thus, we contribute an approach to proactively make the unexpected use and recombination of data visible and normatively assessable, as highlighted by Jarvenpaa and Essén (2023). As such, we contribute to the regulatory demand regarding data ecosystems (Lis et al., 2023).

While existing privacy models generalize the actors involved in personal data sharing (Benson et al., 2015; Conger et al., 2013), AT based on our meta-model not only allows a flexible classification of actors but also captures the data flows between them by relating socio-technical elements across different architectures (EA, IA, and BA). Owing to the generalizability of our results, achieved by involving several experts and heterogeneous cases, AT can be applied to case study research in the privacy field and support the in-depth analysis and comparison of privacy violations. For example, considering selected concerns in a cross-case analysis can reveal patterns of socio-technical relations commonly leading to privacy violations (e.g., technical backdoors via SDKs). Moreover, modeling data ecosystems can show gaps in data flows or indicate whether actors purposely obfuscated parts (e.g., unclear data sources of a processing activity). Thus, from a regulatory perspective, it becomes clear which questions are open and which actors must be scrutinized. By enabling the identification of patterns and gaps in data ecosystems, AT complements analytical frameworks within the privacy field.

Based on our study, we provide a research framework for AT about data ecosystems (see Fig. 10). This framework encompasses the IA, BA, and EA in data ecosystems. In line with these architectural delineations, the framework highlights the data ecosystem, business, legal, data, and IT landscape layers. Each of these layers plays a distinct role, affecting the overall data ecosystem. Prior research has predominantly concentrated on the business and IT landscape layers of EA, focusing on optimizing business processes, IS implementation, and organizational alignment (Gampfer et al., 2018; Kotusev, 2018; Saat et al., 2010; Simon et al., 2013).

The framework emphasizes the potential for innovative research on data ecosystems by highlighting the complex connections between architectural layers. Studying these layers individually and their interconnections is crucial for developing a comprehensive understanding of data ecosystems and underscores the importance of inquiry in advancing knowledge on data ecosystem architectures. Considering different perspectives, such as inter-organizational versus customer-centric, may be particularly valuable.

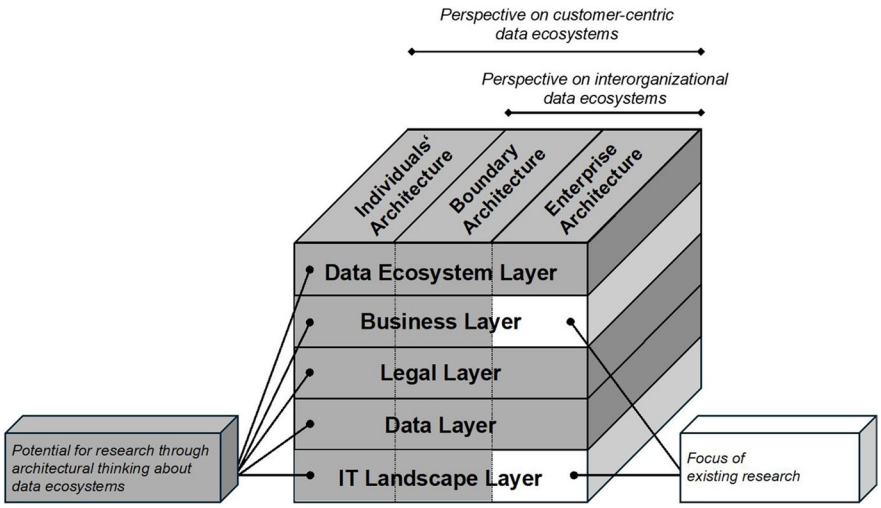

Fig. 10 Research framework for architectural thinking about data ecosystems

Practical implications

Our results have several practical implications. Considering Knackstedt et al.'s (2014) research agenda for conceptual modeling in law, we deliberately distinguished between business and regulatory perspectives on data ecosystems, leveraging AT for both the application and creation of law. For business stakeholders, focusing on the application of law, AT guided by our meta-model can support compliance efforts. For example, modeling data ecosystems can help businesses maintain records of processing activities (GDPR, 2016, Art. 30), identify privacy risks and improve related measures, and proactively demonstrate compliance, e.g., through simplified architecture visualizations. AT can also support business stakeholders in optimizing personal data flows for business processes and in designing or joining a data ecosystem considering privacy constraints. Businesses can rely on the meta-model to recall privacy-critical interdependencies and foresee challenges of data sharing.

For regulatory stakeholders involved in the creation of law, AT helps scrutinize socio-technical relations causing privacy violations in data ecosystems. It also supports anticipating adverse consequences when balancing legal obligations with the benefits of personal data processing. AT further enables the simulation of non-compliance scenarios and, drawing on architectural insights, informs the revision of regulations or privacy statements to strengthen data protection and improve compliance.

Business and regulatory stakeholders can perform AT, guided by the concerns and meta-model, to gain transparency about data ecosystems at different levels of detail during analysis (as-is) and planning (to-be) scenarios. Furthermore, practitioners can extend the identified privacy concerns and tailor the meta-model to their specific needs. We acknowledge that our research focuses on the GDPR as an exemplary regulation. However, the abstract and flexible nature of our meta-model makes it adaptable to other data protection regulations (e.g., CCPA), though practical adjustments by adding, modifying, or removing elements may be necessary. Through its holistic view of data ecosystems, AT based on our meta-model complements methods and tools practitioners use to detect and document privacy challenges, such as checklists or evaluation procedures.

Limitations and future research

Our study is not without limitations. First, we acknowledge the potential bias inherent in the selection of experts and cases. To mitigate this bias, we grounded our results using a mixed-methods approach that combined public data on privacy scandals with 14 expert interviews and evaluated the findings through a multiple case study, balancing diversity and relevance according to theoretical replication logic. However, additional case studies and interviews could reveal other relevant privacy concerns, further refine our meta-model, and demonstrate its application across additional domains, thereby strengthening the generalizability of our findings. For example, surveys with individuals could provide deeper insights to refine the elements of the IA.

Second, our research explored privacy in data ecosystems from a broad perspective. On the one hand, we argue that our meta-model can cover certain data intermediaries (Möller et al., 2024) and data ecosystem types (Curry & Sheth, 2018). For example, the HaMagen data ecosystem can be understood as a data collaborative designed to generate societal benefits, and our modeling effectively revealed privacy issues. On the other hand, future research should identify privacy concerns and socio-technical elements specific to selected data intermediaries and data ecosystem types to enable more fine-granular modeling and analysis of privacy issues. For example, for certain configurations of data spaces, further boundary resources (Eaton et al., 2015; Otto & Jarke, 2019) could be flexibly added to the meta-model. Moreover, our meta-model may inspire the development of additional meta-models that predefine elements tailored to specific configurations, such as data prosumers, consumers, and providers as actors in data spaces (Möller et al., 2024).

Third, while the flexible, concern-driven nature of our meta-model allows stakeholders to tailor its application to specific contexts, this adaptability can introduce subjectivity into the modeling process. The selection and emphasis of certain elements may vary based on individual perspectives, potentially leading to inconsistent applications across different scenarios. Although this issue aligns with the general limitations of interpretative approaches, future research should explore standardized guidelines, tool-supported procedures, and validation mechanisms to mitigate this challenge. Such solutions would help modelers make informed and balanced selections, ultimately enhancing the consistency and reliability of the meta-model's application.

Fourth, although AT aims to be lightweight to address non-IT experts, we identified numerous concerns and elements requiring a complex meta-model. Our demonstrations have shown that close collaboration between IS and legal experts is necessary to jointly decompose data ecosystems and uncover privacy issues. For example, in the HaMagen study, it became evident that architectural models are entirely understandable to legal experts. However, future research should explore ways to simplify data ecosystem models to improve their visual distinction from meta-models, make them less formalized, and enhance their comprehensibility, e.g., to inform individuals about personal data flows. One possible approach is to conduct practical studies on how Moody's (2009) principles can be better aligned with AT, e.g., to reduce symbol overload (as observed in our models) and enhance cognitive effectiveness.

Finally, our research is limited to the context of privacy. Future research should explore how AT can support other data governance challenges and extend our findings. Moreover, our demonstration presents only a selection of privacy issues (e.g., HaMagen's incomplete privacy policy, Reveal Mobile circumventing opt-out mechanisms) that AT can identify, along with exemplary ways AT supports mitigation (e.g., informing privacy-related design decisions, deriving GDPR recommendations). Additional case studies must explore AT's full potential in unraveling and mitigating further privacy risks, such as data leakage or unauthorized third-party access, across various data ecosystem configurations. For example, in another study, AT helped compare location-based services and identify archetypes of privacy issues, ultimately facilitating mitigation (Burmeister et al., 2021).

Conclusion

In an era marked by opaque personal data processing and sharing within complex data ecosystems, business and regulatory stakeholders are challenged to maintain transparency about the various data flows between actors and their processing activities (Crain, 2018; Elrick, 2021; Kurtz et al., 2018). This transparency stands as an essential counterbalance to the benefits of personal data processing, allowing stakeholders to navigate privacy risks from individual and organizational perspectives while comprehending the manifold triggers for privacy violations, thereby enabling regulatory adjustments (Nissenbaum, 2011; Tene & Polonetsky, 2013). We propose AT as a promising paradigm for attaining transparency, facilitating the systematic decomposition and analysis of data ecosystems. By identifying key privacy concerns of business and regulatory stakeholders and developing a corresponding data ecosystem architecture meta-model, this study lays foundational groundwork for leveraging AT in the privacy context. Researchers and practitioners can use our results to unravel dependencies in data ecosystems through a socio-technical lens, identify privacy-critical issues across intra- and inter-organizational boundaries, and facilitate privacy-related endeavors.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Aaen, J., Nielsen, J. A., & Carugati, A. (2022). The dark side of data ecosystems: A longitudinal study of the DAMD project. *European Journal of Information Systems, 31*(3), 288–312. https://doi.org/10.1080/0960085X.2021.1947753

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature, 54*(2), 442–492. https://doi.org/10.1257/jel.54.2.442

Aier, S., Labusch, N., & Pähler, P. (2015). Implementing architectural thinking: A case study at Commerzbank AG. In A. Persson & J. Stirna (Eds.), *Trends in enterprise architecture research – CAiSE 2015 International Workshops* (pp. 389–400). Springer. https://doi.org/10.1007/978-3-319-19243-7_36

Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society, 20*(3), 973–989. https://doi.org/10.1177/1461444816676645

Attard, J., Orlandi, F., & Auer, S. (2016). Data value networks: Enabling a new data ecosystem. In *Proceedings of the 2016 International Conference on Web Intelligence* (pp. 453–456). https://doi.org/10.1109/WI.2016.0073

Azkan, C., Möller, F., Meisel, L., & Otto, B. (2020). Service dominant logic perspective on data ecosystems - a case study based morphology. *Proceedings of the 28th European Conference on Information Systems, virtual.*

Azkan, C., Möller, F., Ebel, M., Iqbal, T., Otto, B., & Poeppelbuss, J. (2022). Hunting the treasure: Modeling data ecosystem value co-creation. *Proceedings of the 43rd International Conference on Information Systems*.

Bélanger, F. (2012). Theorizing in information systems research using focus groups. *Australasian Journal of Information Systems, 17*(2), 109–135. https://doi.org/10.3127/ajis.v17i2.695

Belo, Í., & Alves, C. (2021). How to create a software ecosystem? A partnership meta-model and strategic patterns. *Information, 12*(6), 1–29. https://doi.org/10.3390/info12060240

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly, 11*(3), 369–386. https://doi.org/10.2307/248684

Benson, V., Saridakis, G., & Tennakoon, H. (2015). Information disclosure of social media users: Does control over personal information, user awareness and security notices matter? *Information Technology & People, 28*(3), 426–441. https://doi.org/10.1108/ITP-10-2014-0232

Burmeister, F., Drews, P., & Schirmer, I. (2019a). An ecosystem architecture meta-model for supporting ultra-large scale digital transformations. In *Proceedings of the 25th Americas Conference on Information Systems*.

Burmeister, F., Drews, P., & Schirmer, I. (2019b). Leveraging architectural thinking for large-scale e-government projects. In *Proceedings of the 40th International Conference on Information Systems*.

Burmeister, F., Drews, P., & Schirmer, I. (2019c). A privacy-driven enterprise architecture meta-model for supporting compliance with the General Data Protection Regulation. In *Proceedings of the 52nd Hawaii International Conference on System Sciences* (pp. 6052–6061).

Burmeister, F., Drews, P., & Schirmer, I. (2021). Modeling the c(o)urse of privacy-critical location-based services – Exposing dark side archetypes of location tracking. In *Proceedings of the 54th Hawaii*

International Conference on System Sciences (pp. 6651–6660). Virtual.

Burmeister, F., Zar, M., Böhmann, T., Elkin-Koren, N., Kurtz, C., & Schulz, W. (2022). Toward architecture-driven interdisciplinary research – Learnings from a case study of COVID-19 contact tracing apps. In *Proceedings of the 2nd ACM Symposium on Computer Science and Law* (pp. 143–154). https://doi.org/10.1145/3511265.3550451

Cate, F. H., & Mayer-Schönberger, V. (2013). Notice and consent in a world of big data. *International Data Privacy Law, 3*(2), 67–73. https://doi.org/10.1093/idpl/ipt005

Conger, S., Pratt, J. H., & Loch, K. D. (2013). Personal information privacy and emerging technologies. *Information Systems Journal, 23*(5), 401–417. https://doi.org/10.1111/j.1365-2575.2012.00402.x

Crain, M. (2018). The limits of transparency: Data brokers and commodification. *New Media & Society, 20*(1), 88–104. https://doi.org/10.1177/1461444816657096

Culnan, M. J., & Bies, R. J. (2003). Consumer privacy: Balancing economic and justice considerations. *Journal of Social Issues, 59*(2), 323–342. https://doi.org/10.1111/1540-4560.00067

Curry, E., & Sheth, A. (2018). Next-generation smart environments: From system of systems to data ecosystems. *IEEE Intelligent Systems, 33*(3), 69–76. https://doi.org/10.1109/MIS.2018.033001418

Da Silva, A. R. (2015). Model-driven engineering: A survey supported by the unified conceptual model. *Computer Languages, Systems & Structures, 43*, 139–155. https://doi.org/10.1016/j.cl.2015.06.001

Demchenko, Y., de Laat, C., & Membrey, P. (2014). Defining architecture components of the big data ecosystem. In *Proceedings of the 2014 International Conference on Collaboration Technologies and Systems* (pp. 104–112). https://doi.org/10.1109/CTS.2014.6867550

Eaton, B., Elaluf-Calderwood, S., Sorensen, C., & Yoo, Y. (2015). Distributed tuning of boundary resources: The case of Apple's iOS service system. *MIS Quarterly, 39*(1), 217–243. https://doi.org/10.25300/MISQ/2015/39.1.10

Elrick, L. E. (2021). The ecosystem concept: A holistic approach to privacy protection. *International Review of Law, Computers & Technology, 35*(1), 24–45. https://doi.org/10.1080/13600869.2020.1784564

Erraissi, A., & Belangour, A. (2018). Data sources and ingestion big data layers: Meta-modeling of key concepts and features. *International Journal of Engineering & Technology, 7*(4), 3607–3612. https://doi.org/10.2139/ssrn.3185342

Flick, U. (2009). *An introduction to qualitative research*. Sage.

Frank, U. (2014). Multi-perspective enterprise modeling: Foundational concepts, prospects and future research challenges. *Software & Systems Modeling, 13*(3), 941–962. https://doi.org/10.1007/s10270-012-0273-9

Gampfer, F., Jürgens, A., Müller, M., & Buchkremer, R. (2018). Past, current and future trends in enterprise architecture – a view beyond the horizon. *Computers in Industry, 100*, 70–84. https://doi.org/10.1016/j.compind.2018.03.006

GDPR. (2016). General data protection regulation.Regulation (EU) 2016/679 of the European Parliament and of the Council, *Official Journal of the European Union* (111).

Gelhaar, J., Groß, T., & Otto, B. (2021). A taxonomy for data ecosystems. In *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS), a virtual conference* (pp. 6113–6122). https://doi.org/10.24251/HICSS.2021.739

Gopal, R. D., Hidaji, H., Patterson, R. A., Rolland, E., & Zhdanov, D. (2018). How much to share with third parties? User privacy concerns and website dilemmas. *MIS Quarterly, 42*(1), 143–164. https://doi.org/10.25300/MISQ/2018/13839

Heinz, D., Benz, C., Fassnacht, M., & Satzger, G. (2022). Past, present and future of data ecosystems research: A systematic literature review. *Proceedings of the Pacific Asia Conference on Information Systems*, virtual.

Horlach, B., Drechsler, A., Schirmer, I., & Drews, P. (2020). Everyone's going to be an architect: Design principles for architectural thinking in agile organizations. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 6197–6206). https://doi.org/10.24251/HICSS.2020.759

IEEE. (2003). IEEE Std 1175.1–2002. IEEE Guide for CASE Tool Interconnections – Classification and Description.

ISO/IEC. (2002). Information technology - CDIF framework - Part 1: Overview. Standard 15474–1:2002.

ISO/IEC/IEEE.(2011). Systems and software engineering – architecture description. Standard 42010:2011.

Jarvenpaa, S. L., & Essén, A. (2023). Data sustainability: Data governance in data infrastructures across technological and human generations. *Information and Organization, 33*(1), 100449. https://doi.org/10.1016/j.infoandorg.2023.100449

Karwatzki, S., Trenz, M., Tuunainen, V. K., & Veit, D. (2017). Adverse consequences of access to individuals' information: An analysis of perceptions and the scope of organisational influence. *European Journal of Information Systems, 26*(6), 688–715. https://doi.org/10.1057/s41303-017-0064-z

Knackstedt, R., Heddier, M., & Becker, J. (2014). Conceptual modeling in law: An interdisciplinary research agenda. *Communications of the Association for Information Systems, 34*(36), 711–736. https://doi.org/10.17705/1CAIS.03436

Kotusev, S. (2018). *The practice of enterprise architecture: A modern approach to business and IT alignment*. SK Publishing.

Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2017). The (unfulfilled) potential of data marketplaces. *ETLA working papers*, no. 53.

Kühne, T. (2006). Matters of (meta-) modeling. *Software and Systems Modeling, 5*(4), 369–385. https://doi.org/10.1007/s10270-006-0017-9

Kurtz, C., Semmann, M., & Schulz, W. (2018). Towards a framework for information privacy in complex service ecosystems. In *Proceedings of the 39th International Conference on Information Systems*.

Kurtz, C., & Burmeister, F. (2024). Multi-role actors and rebounding effects across user interfaces - Exploring Big Tech's privacy scandals and GDPR limitations in data ecosystems. In A. Marcus, E. Rosenzweig, M. M. Soares, P.-L. P. Rau & A. Moallem (Eds.), *HCI International 2024 – Late Breaking Papers. Lecture Notes in Computer Science*, (Vol. 15380, pp. 283–303). Springer. https://doi.org/10.1007/978-3-031-76821-7_20

Lagerström, R., Saat, J., Franke, U., Aier, S., & Ekstedt, M. (2009). Enterprise meta-modeling methods – combining a stakeholder-oriented and a causality-based approach. In T. A. Halpin, J. Krogstie, S. Nurcan, E. Proper, R. Schmidt, P. Soffer, & R. Ukor (Eds.), *Business-process and information systems modeling, LNBIP 29* (pp. 381–393). Springer. https://doi.org/10.1007/978-3-642-01862-6_31

Li, H., Yu, L., & He, W. (2019). The impact of GDPR on global technology development. *Journal of Global Information Technology Management, 22*(1), 1–6. https://doi.org/10.1080/1097198X.2019.1569186

Lis, D., Gelhaar, J., & Otto, B. (2023). Data strategy and policies: The role of data governance in data ecosystems. In I. Caballero & M. Piattini (Eds.), *Data governance: From the fundamentals to real cases* (pp. 27–55). Springer. https://doi.org/10.1007/978-3-031-43773-1_2

Martin, N., Matt, C., Niebel, C., & Blind, K. (2019). How data protection regulation affects startup innovation. *Information Systems Frontiers, 21*, 1307–1324. https://doi.org/10.1007/s10796-019-09974-2

Mayring, P. (2014). *Qualitative content analysis: Theoretical foundation, basic procedures and software solution.* Beltz.

Möller, F., Jussen, I., Springer, V., Gieß, A., Schweihoff, J. C., Gelhaar, J., Guggenberger, T., & Otto, B. (2024). Industrial data ecosystems and data spaces. *Electronic Markets, 34*, 41. https://doi.org/10.1007/s12525-024-00724-0

Moody, D. L. (2009). The "physics" of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Transactions on Software Engineering, 35*(6), 756–779. https://doi.org/10.1109/TSE.2009.67

Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and Organization, 17*(1), 2–26. https://doi.org/10.1016/j.infoandorg.2006.11.00

Myers West, S. (2019). Data capitalism: Redefining the logics of surveillance and privacy. *Business & Society, 58*(1), 20–41. https://doi.org/10.1177/0007650317718185

Niemi, E. (2007). Enterprise architecture stakeholders - a holistic view.*Proceedings of the 13th Americas Conference on Information Systems*, Keystone, CO, USA.

Nissenbaum, H. (2011). A contextual approach to privacy online. *Daedalus, 140*(4), 32–48. https://doi.org/10.1162/DAED_a_00113

Nissenbaum, H. (2019). Contextual integrity up and down the data food chain. *Theoretical Inquiries in Law, 20*(1), 221–256. https://doi.org/10.1515/til-2019-0008

Object Management Group. (2002). Meta Object Facility (MOF) Specification. *Version, 1*, 4.

Oliveira, M. I. S., Oliveira, L. E. R. A., Batista, M. G. R., & Lóscio, B. F. (2018). Towards a meta-model for data ecosystems. In *Proceedings of the 19th Annual International Conference on Digital Government Research*. https://doi.org/10.1145/3209281.3209333

Oliveira, M. I. S., de Fátima Barros Lima, G., & Lóscio, B. F. (2019). Investigations into data ecosystems: A systematic mapping study. *Knowledge and Information Systems, 61*(2), 589–630. https://doi.org/10.1007/s10115-018-1323-6

Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the international data spaces case. *Electronic Markets, 29*(4), 561–580. https://doi.org/10.1007/s12525-019-00362-x

Pidd, M. (2009). *Tools for thinking: Modelling in management science*. Wiley.

Purtova, N. (2018). The law of everything. Broad concept of personal data and future of EU data protection law. *Law, Innovation and Technology, 10*(1), 40–81. https://doi.org/10.1080/17579961.2018.1452176

Saat, J., Franke, U., Lagerström, R., & Ekstedt, M. (2010). Enterprise architecture meta models for IT/business alignment situations. In *Proceedings of the 14th IEEE EDOC* (pp. 14–23). https://doi.org/10.1109/EDOC.2010.17

Saldaña, J. (2015). *The coding manual for qualitative researchers*. Sage.

Sandkuhl, K., Fill, H.-G., Hoppenbrouwers, S., Krogstie, J., Matthes, F., Opdahl, A., Schwabe, G., Uludag, Ö., & Winter, R. (2018). From expert discipline to common practice: A vision and research agenda for extending the reach of enterprise modeling. *Business & Information Systems Engineering, 60*(1), 69–80. https://doi.org/10.1007/s12599-017-0516-y

Scheider, S., Lauf, F., Möller, F., & Otto, B. (2023). A reference system architecture with data sovereignty for human-centric data ecosystems. *Business & Information Systems Engineering, 65*(5), 577–595. https://doi.org/10.1007/s12599-023-00816-9

Simon, D., Fischbach, K., & Schoder, D. (2013). An exploration of enterprise architecture research. *Communications of the Association for Information Systems, 32*(1), 1–72. https://doi.org/10.17705/1CAIS.03201

Stachowiak, H. (1973). *Allgemeine Modelltheorie*. Springer.

Tene, O., & Polonetsky, J. (2013). Big data for all: Privacy and user control in the age of analytics. *Northwestern Journal of Technology and Intellectual Property, 11*(5), 239–273.

The Open Group. (2018). TOGAF standard, version 9.2.

Winter, R. (2014). Architectural thinking. *Business & Information Systems Engineering, 6*(6), 361–364. https://doi.org/10.1007/s11576-014-0439-x

Yin, R. K. (2009). *Case study research: Design and methods*. Sage.