A Privacy-driven Enterprise Architecture Meta-Model for Supporting Compliance with the General Data Protection Regulation

Fabian Burmeister
University of Hamburg
burmeister@informatik.uni-hamburg.de

Paul Drews
Leuphana University of Lüneburg
paul.drews@leuphana.de

Ingrid Schirmer
University of Hamburg
schirmer@informatik.uni-hamburg.de

Abstract

The processing of personal data has evolved into an integral component of businesses by providing several data-driven opportunities. Simultaneously, businesses struggle with the associated responsibility for privacy, as recent data scandals have shown. As a consequence, the European Commission has passed the General Data Protection Regulation (GDPR) to enhance the rights of citizens and the requirements on data protection. This paper argues that enterprise architecture (EA) models can be a key to compliance with the GDPR. Following an incremental research approach, we categorize the major obligations resulting from the GDPR, derive essential stakeholder concerns and outline necessary EA elements for capturing aspects of analytics, security and privacy in EA models. On this basis, a privacy-driven EA meta-model is developed that is capable of answering key concerns resulting from the GDPR.

1. Introduction

Big data has rapidly become the revolutionizer of our digital world. Recent advances in data mining, complex algorithms and artificial intelligence have led to significant breakthroughs by processing and analyzing large data sets. Being the fuel of the 21st century, data have become the new source of enormous economic and social value, causing a shift from physical product development towards information aggregation [1]. The fastest-growing companies in history are those, who rely on data-driven business models: Alibaba, the world's most valuable retailer, has no inventory; Facebook, the most popular media owner, creates no content; Uber, the largest taxi company, has no own vehicles [1].

However, big data's role as a value creator comes along with a dark side. While businesses are forced to process and analyze data to understand their changing customer needs and withstand competition, they are also required to provide innovative data-driven services and products [2, 3]. At the same time, the data deluge is often composed of personal information, gleaned from a wealth of heterogeneous sources like social media, online transactions, health records, global positioning and physical sensors, raising privacy concerns that could trigger a regulatory backlash, dampen the data economy and stifle innovation [3, 4]. Previous and recent privacy scandals and data breaches, such as the Facebook-Cambridge Analytica scandal, illuminate that awareness of privacy is playing an increasingly crucial role.

According to the well-known definition of Alan Westin, privacy is "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" [5]. Considering the number of privacy scandals and the amount of personal data that is being collected, processed and shared with or without the individual's explicit knowledge, it is obvious that this claim is not completely fulfilled today. In addition, privacy laws and data protection regulations vary greatly between countries. Therefore, in April 2016, the European Commission has passed the General Data Protection Regulation (GDPR) to address the issues of privacy by a unified regulation and balance between beneficial use of personal data and the protection of individual privacy.

Since May 2018, enterprises have to comply with the GDPR and its set of 99 articles, otherwise they can be fined up to four percent of their revenue [6, Art. 83]. This underlines the importance for enterprises of being completely aware of their legal requirements, having transparency about their storing, processing and sharing of personal data and understanding the associated relationships along their whole enterprise architecture (EA). Moreover, enterprises are forced to implement appropriate organizational and technical measures to guarantee security, inform about their use of personal data and adapt their big data analytics processes to attain full compliance with the GDPR [7, 8]. Security, in this context, can be defined as the means for protecting data by ensuring their confidentiality, availability and integrity [9]. To ensure GDPR compliance and support continuous transformation driven by big data analytics, enterprises demand for models that illustrate both the

URI: https://hdl.handle.net/10125/60040
ISBN: 978-0-9981331-2-6

use and security of personal data [7, 8]. With the aim of achieving transparency, consistency and measurability of business and IT components [10], EA modeling provides a reasonable approach for this challenge.

Against this background, we aim to 1) study the state of the art on integrating EA, security and privacy, 2) identify relevant stakeholder concerns and EA elements by analyzing the GDPR, 3) develop a privacy-driven EA meta-model to support processing and protection of personal data, and 4) discuss implications from the meta-model for the enterprise architecture management (EAM), which aims to constantly align business and IT [9]. Therefore, we apply to the following research question: *Which elements and relations need to be included in an EA meta-model for addressing GDPR-related stakeholder concerns?* By answering the given research question, we aim to contribute to current research on the interplay between EA, security and privacy and aspire to provide useful value for practice.

In the following section, we summarize related research. Section 3 outlines our research approach. Section 4 describes four categories of privacy-related obligations that we identified by analyzing the GDPR. Section 5 presents derived stakeholder concerns and EA elements. In section 6, we present our privacy-driven EA meta-model. Section 7 discusses implications from the meta-model for the EAM. Finally, the paper closes with a summary and an outlook.

2. Related research

Over the last few decades, both research and practice have developed many meta-models for describing the layers, artifacts and attributes of EA, shifting EA modeling to a well-researched field [11]. By capturing as-is and to-be models of their EA, organizations aim to optimize business IT alignment, receive transparency about current and future states and realize architectural transitions through EAM as smoothly as possible. For deriving their specific EA models, enterprises rely on EA meta-models as they ensure "semantic rigor, interoperability and traceability" [12]. Since the use of personal data in context of big data analytics and compliance with the GDPR affect the whole EA, innovative EA meta-models are required that address the increasing challenges resulting from privacy [8].

In the literature, EA meta-models, if any, merely capture the protection of personal data as a superior issue that needs to be managed, but do not particularize privacy- and security-related artifacts and attributes and their interrelations with existing EA elements [9, 10, 12]. Consequently, privacy and security architectures are often still separated from the EA [13, 14]. The Open Group states: "For too long, information security has been considered a separate discipline, isolated from the enterprise architecture" [13]. For this reason, some streams of research attempt to complement the EA with privacy- or security-related aspects by developing an enterprise privacy architecture (EPA) and enterprise security architecture (ESA). Nevertheless, a direct integration of privacy- and security-relevant aspects is still missing, since these approaches provide additional architectures alongside the existing EA. As a result, several gaps are occurring between the architectural perspectives, which need to be overcome [13]. In the following, we briefly summarize the approaches that aim for bringing together privacy, security and EA.

2.1. Enterprise privacy architecture

There is neither a standard definition of an EPA nor a homogenization of its granularity, structure or components in the literature. A relatively well-known representative, however, is the IBM EPA, which defines itself as "a methodology that allows enterprises to maximize the business use of personal information while respecting privacy concerns and regulations" [15]. It contains a modular structure, consisting of four building blocks: A privacy regulation analysis for identifying applicable regulations, a management reference model for defining the strategy, controls and practices for privacy in an enterprise, a privacy agreements framework that models privacy-relevant players, data and rules to enable a privacy-enhanced business process reengineering and finally a technical reference architecture that defines the technology for implementing required privacy services [15]. Although the IBM EPA provides essential building blocks for ensuring privacy, it rather embodies a general guideline instead of a concrete meta-model and therefore does not illustrate relations to existing elements of EA.

2.2. Enterprise security architecture

Without having an adequate security management, enterprises cannot guarantee privacy. Nevertheless, existing security technologies and services often provide security, but not privacy [15]. For instance, non-anonymous identification and authentication schemes, data collected by intrusion detection systems and coarse access control [15]. In order to ensure privacy, these security measures require a transformation into privacy-enabling security services by an integration into the whole EA. Hence, an ESA seeks to translate a vision of information security into effective enterprise evolution by capturing a current and future state of an enterprise's security controls, including policies, security processes, information security systems and organizational units, so that they align with strategic goals and business

objectives [16, 17]. In contrast to the EPA, the approach of an ESA is actually more popular in both science and practice. Gartner for instance, being inspired by EA frameworks, recommends three levels of abstraction (conceptual, logical and implementation) and three related viewpoints (business, information and technical) within an ESA [17]. Another approach is the Sherwood Applied Business Security Architecture (SABSA), which consists of five horizontal layers (contextual, conceptual, logical, physical and component) and one vertical layer (operational) for realizing security services [16]. Compared to Gartner's approach, which is more theoretical, SABSA comes along with a more practical oriented methodology [17]. A third attempt towards an ESA is the Open Security Architecture (OSA), which provides a complex library of patterns, controls and threats for ensuring security [18]. However, in contrast to the aforementioned, the OSA does not embody a concrete framework, but provides a detailed catalog for security-related assistance. Oda et al. [16] and Shariati et al. [17] compare some additional approaches towards an ESA and provide a decent overview. In summary, they state that more research on the interoperability of ESA is required and underline the increasing importance of connecting key stakeholders from business, information, technology and security layers. They also highlight the need for a closer integration of ESA and EA, since their studies revealed that business and IT components are often developed separately from security components.

2.3. Enterprise architecture and GDPR

Although the GDPR was passed two years ago, there is hardly any scientific publication on the interplay between GDPR and EA. Accordingly, we consulted blogs, technical reports and white papers of EA tool providers in order to get an overview of the current state of the art. Lankhorst [19], for instance, underlines that the GDPR not only demands compliance, but also requires a concrete demonstration of compliance. He accentuates that EA models are a major source of information, since they could provide a "coherent and connected view of everything related to personal data" [19]. Other sources additionally highlight that EA modeling is an enabler of privacy by design, as claimed by the GDPR [6, Art. 25], because it gives transparency about interconnections of an organization's systems and therefore about the data flows along the application development lifecycle [8, 20]. EA tool providers also state that the role of enterprise architects as an essential interface to numerous stakeholders, particularly the data protection officer [6, Art. 37], is becoming even more important by being able to answer GDPR-related concerns by EA models [19, 20]. Additionally, we found out that many users of EA tools still have to rely on custom workarounds for modeling the processing and security of personal data, because they lack a consistent approach towards the topic as well as privacy-relevant artifacts and attributes [7, 19]. Therefore, it is our ambition to derive an EA meta-model that includes essential insights from EPA and ESA on the one hand and delivers guidance by specific modeling elements for supporting GDPR compliance on the other hand.

3. Research approach

By analyzing literature about the interplay of EA, security and privacy as described in the previous section, we found out that an EA meta-model focusing on the processing of personal data is missing so far. Especially a consideration of privacy-relevant elements according to the GDPR and the integration of a security architecture within EA meta-models would create additional value for both research and practice [14]. To address this research gap and develop our EA meta-model, we adopted a design science oriented multi-methodological research approach consisting of three consecutive steps (see Figure 1). We followed a top-down conceptual analysis based on stakeholder concerns as described in [21] to concretize the information needs resulting from the GDPR first and derive EA elements afterwards. In our context of research, we define stakeholders as individuals that aim to achieve compliance with the GDPR.

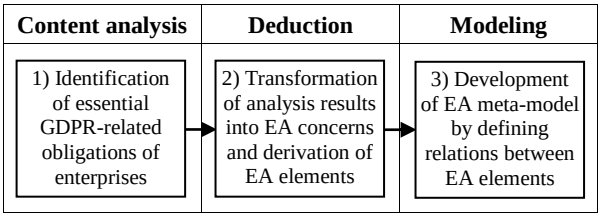

Figure 1. Research approach

During the first step, we conducted a structured in-depth content analysis of the GDPR to identify the major obligations for enterprises by following the procedure proposed in [22]. According to the GDPR, enterprises can play both the role of a controller and a processor, which have specific obligations. While the controller determines the purpose of processing [6, Art. 4 (7)], the processor realizes and executes appropriate analytical procedures to process personal data on behalf of the controller [6, Art. 4 (8)]. Moreover, the controller has to comply with the manifold rights of the data subject, an identifiable natural person whose personal data are processed [6, Art. 4 (1)], and to report to a supervisory authority that monitors compliance [6, Art. 51]. Hence, we defined the direction of our content analysis as the essential paragraphs that an enterprise has to fulfill. By coding the analyzed content, we grouped the results into four categories of major obligations (section 4).

In the second step, we referred to the previously identified obligations and deductively derived relevant EA concerns to concretize the requirements of the GDPR. Under consideration of elements and relations proposed by existing EA meta-models and referring to the literature identified in section 2, we derived privacy-related EA artifacts and attributes for addressing these concerns (section 5).

Finally, in a third step, we developed the privacy-driven EA meta-model by relating the EA elements and arranging the layers and attributes. By referring back to selected concerns of each category, we demonstrated the EA meta-model and discussed implications for the EAM (sections 6 and 7).

4. GDPR-related obligations of enterprises

In the following, we summarize our insights about the essential obligations of enterprises according to the GDPR. This analysis serves as our basis to deductively derive and discuss relevant EA concerns and elements afterwards. The following Figure 2 gives an overview of our determined categorization of the obligations:

Category A Compliance with superior principles (Art. 5, 6, 7, 8)	**Category B** Information obligations (Art. 12, 13, 14, 19, 30, 33)
Category C Satisfaction of data subject's rights (Art. 15, 16, 17, 18, 20, 21)	**Category D** Implementation and verification of organizational and technical measures (Art. 24, 25, 28, 32, 35, 37, 38, 39)

Figure 2. Matrix of GDPR-related obligations of enterprises

In our categorization, we only refer to the articles that directly have a great impact on the controller or processor entity. The remaining articles are not included, since they describe severability clauses, focus on the interplay and behavior of other entities, such as the supervisory authority and European data protection board, or specify miscellaneous aspects related to remedies, liability and penalties.

4.1. Compliance with superior principles

Enterprises are required to process all personal data in a lawful, fair and transparent manner [6, Art. 5 (1)] and to collect them only for specified, explicit and legitimate purposes [6, Art. 5 (2)]. In addition, the processing of personal data should follow the principle of data minimization or rather be limited to what is necessary for achieving defined purposes [6, Art. 5 (3)]. Moreover, personal data shall be accurate and, where indispensable, kept up to date [6, Art. 5 (4)], only be stored as long as necessary [6, Art. 5 (5)] and be processed in a manner that ensures appropriate security [6, Art. 5 (6)]. The processing and use of personal data is only allowed if an enterprise has a traceable permission [6, Art. 6, Art. 7, Art. 8], which may arise from the GDPR itself (e.g., the processing is necessary to fulfill a contract) or through the explicit consent of the data subject. To comply with the above-mentioned principles, enterprises have to justify and document the exact purpose of storing specific data. They are also required to recognize where and how long which data are stored in order to guarantee deletion or updates. This results in a big challenge, since many enterprises kept obsolete personal data for possible future purposes.

4.2. Information obligations

Enterprises have several information obligations to both the data subject and the supervisory authority. Generally, involved parties have to be informed in a concise, transparent, intelligible and easily accessible form [6, Art. 12]. Enterprises in the function of a controller have to provide several pieces of information to the data subject, primarily the purpose of processing, the duration of data storage, the sources of collected personal data and in case of an automated decision-making, including profiling, meaningful information about the involved logic [6, Art. 13, Art. 14]. Moreover, the controller shall notify recipients of personal data about a rectification or deletion of personal data as well as a restriction of processing [6, Art. 19]. In addition, both the controller and the processor have to make a record of processing activities available to the responsible supervisory authority [6, Art. 30]. That record should include, but is not limited to, the purpose of processing, affected data subjects and recipients, the categories of personal data, the intended time limits for an erasure of data categories, transfers of personal data to a third country as well as a general description of technical and organizational security measures [6, Art. 30]. Above all, enterprises are obliged to notify data breaches to the supervisory authority and concerned data subjects within 72 hours of being aware [6, Art. 33]. This notification shall contain a description of the nature of the data breach including the approximate number of concerned data subjects and data records, an estimation of possible consequences and a statement of measures taken or proposed [6, Art. 33]. Especially the record of processing activities and the need to notify data breaches on time force enterprises to be completely aware of their use and security of personal data, resulting in a high demand for constant transparency and documentation.

4.3. Satisfaction of data subject's rights

The data subject has a multitude of rights that confront the controller with some challenges in handling personal data. First, when requested by the data subject, the controller has to provide a copy of the personal data undergoing processing ("Right of access" [6, Art. 15]). Additionally, the data subject can demand an immediate correction and completion of personal data ("Right of rectification" [6, Art. 16]) and that specific personal data may be deleted without delay ("Right to be forgotten" [6, Art. 17]). The "right of restriction" in addition, forces the controller to limit the processing of a data subject's personal data under certain conditions [6, Art. 18]. Controllers also have to provide personal data to the concerned data subject in a structured, commonly used and machine-readable format and, where technically feasible, transmit the personal data electronically to another controller ("Right of data portability" [6, Art. 20]). Finally, enterprises have to refrain from processing personal data, if a data subject files an objection ("Right to object" [6, Art. 21]). Fulfilling these rights requires not only a complete tracking of personal data within an enterprise, but also an understanding of the different data formats and the ability to unify these.

4.4. Implementation and verification of organizational and technical measures

For a lawful processing of personal data according to the GDPR, enterprises have to arrange appropriate organizational and technical measures to realize data protection and information security. Referring to [6, Art. 24, Art. 28], both the controller and processor have to implement, prove and update these measures to ensure and demonstrate compliance with the GDPR. As stated in [6, Art. 25], these measures should also be designed in an effective manner to enforce the privacy principles listed in category A ("privacy by design"), such as data minimization, and that by default, only personal data that are actually necessary for a specific purpose are processed ("privacy by default"). [6, Art. 32] details that technical and organizational measures should include a pseudonymization and encryption of personal data, the ability to ensure an ongoing confidentiality, integrity, availability and resilience of processing systems and services, a rapid recovery of personal data and a process for regular evaluation. Where a type of processing, such as the utilization of new technologies, is likely to entail a high risk for the privacy of data subjects, a privacy impact assessment should be carried out [6, Art. 35]. In addition to privacy by design, this can lead to necessary improvements of the measures. Due to the complexity, the advice of a designated data protection officer should be sought, who monitors and ensures compliance with both an enterprise's security strategy and the GDPR [6, Art. 37, Art. 38, Art. 39]. The obligations of this category, such as privacy by design, force enterprises to constantly be aware of their overall security maturity. To achieve compliance, enterprises require an in-depth transparency about which security measures protect which business and IT components from which type of potential attack.

4.5. Interim conclusion

In summary, the identified categories of obligations confront enterprises with various challenges relating to the documentation, control and security of processing personal data. Enterprises are required to completely understand their data flows, have an overview of their data sources and recognize the security maturity of their data stores. In contrast, the needed in-depth awareness of privacy offers several opportunities. For instance, enterprises could gain valuable insights about potential data-driven improvements of business processes and services, uncover possibilities for homogenizing data analytics processes and tools and identify options to supersede specific data sources. By complying with the GDPR, enterprises may also receive a certificate that proves their privacy-friendliness [6, Art. 42] and, in turn, can lead to greater customer and partner confidence.

We argue that capturing privacy-relevant aspects in EA models and relating these to existing EA elements supports being compliant with the GDPR and reveals opportunities to generate additional value. Having a look at modern business models and recent data scandals as stated in the beginning, a high transparency and awareness of the data-driven coherence with regard to privacy within the whole EA is more important than ever to survive in times of big data [2, 20].

5. GDPR-related EA concerns and elements

Following the top-down research approach to meta-model definition based on stakeholder concerns [21], we refer to the four categories of obligations as our main source for deduction. From each category, we select representative issues for deriving significant concerns and EA elements. We argue that stakeholder concerns concretize the information needs and consequently disclose which EA elements are necessary. In order to substantiate and integrate required EA elements, we additionally fall back on related literature described in section 2 and implement EA elements of existing EA meta-models [9, 12, 13]. For proposing analytics- and security-related elements, we especially refer to [2, 4, 23], since these papers focus on necessary security measures for big data analytics.

5.1. EA concerns and elements for category A

While transparency, an essential principle according to the GDPR, shall be achieved through EA modeling itself, purpose limitation and data minimization require a rationale why certain personal data are processed. In order to additionally comply with the principles of accuracy and storage limitation, personal data that are deemed as inconsistent or irrelevant should be removed or updated. This causes the need for a regular data review, implying the following concerns:

- Why do we collect certain personal data [8]? Which goals and business objects are affected by processing personal data and how important are these for the overall business [19]?
- Which personal data are effectively used [8]? How can we ensure their completeness and consistency?
- What is the maturity of our data? Is it possible to reduce the stored amount of personal data [20]?

To arrange the derived EA elements, we refer to the EA layers according to [9] and suggest an additional processing layer for modeling analytics-related artifacts in detail [23] (see Table 1), since compliance with the GDPR requires high transparency about the internal processing of personal data. On the data layer, we define a data stack as a collection of data objects, which might consist of both business and personal data [4]. The data stack is periodically updated and processed for a specific purpose that supports applications and business objects.

Table 1. EA elements for category A

EA layer	EA artifact	Attributes
Business	Strategic goal	Priority, success criteria
	Business process	Type (value stream, scenario, workflow, detailed procedure), criticality, frequency
	Business service	Type (traditional, data-driven, product-oriented, supportive), criticality, frequency
Application	Business software	Class (ERP, CRM, DMS, HRM), category (standard, individual), version
Processing	Processing purpose	Description, type (decision support, profiling, clustering), priority
Data	Data object	Class (business data, personal data), content, date of storage
	Data stack	Size, complexity (structured, semi-structured, unstructured), portion of personal data

5.2. EA concerns and elements for category B

In section 4.2, we summarized the information obligations towards the data subject and the supervisory authority. The imposed record of processing activities for instance, requires detailed information on the used categories of personal data, the implemented logic of analytics and integrated data protection services. Moreover, the need for notifying data breaches on time necessitates a constant monitoring of infrastructure elements storing personal data and of data streams that realize a connection to recipients and external data sources. Category B, therefore, implies a heterogeneous mix of several stakeholder concerns:

- Which applications process, analyze, visualize and use personal data? Which methods and algorithms realize the implemented logic of analytics [7, 20]?
- How do we arrange our data? Which categories of data contain personal data? Which categories of data require a particularly sensitive approach [19]?
- How regularly do we monitor elements that store or transmit personal data [20]? Do we share personal data with suppliers? How are sensitive data sent?

To answer the latter concern, we refer to our results in section 2.2 and propose modeling a security layer (see Table 2) that shall bridge the gap between EA and ESA [16]. An essential artifact is the data protection service, which enhances network and data security by constantly monitoring allocated elements and providing security measures like server replication and disaster recovery [23]. Transparency on the implemented analytical logic requires a documented processing layer. The analytics tool triggers processing activities, which in turn follow specific processing methods that define the level of abstraction of analytics [2]. The processing methods, in turn, consist of one or more algorithms that are intended to provide purpose-related valuable results [2].

Table 2. EA elements for category B

EA layer	EA artifact	Attributes
Application	Analytics tool	Class (data discovery, data processing, data exploitation, data interfacing), category (standard, individual), version
Processing	Processing activity	Type (manual, automatic), frequency, average duration
	Processing method	Level of abstraction (descriptive, predictive, prescriptive), type (machine learning, natural language processing, computer vision), reliability
	Algorithm	Type (two-/multi-class classification, regression, anomaly detection), average accuracy
Data	Data category	Type (contact data, biometrical data, financial data, GPS data, lifestyle information, medical data), Rating (normal, sensitive)
	Data stream	Type (push, pull), integrated encryption scheme (AES, RSA, DSA, ECC), frequency, latency
	Database	Size, granularity of access rights, number of encrypted records
Technology	Infrastructure element	Class (server, cloud, network, device, sensor), criticality
Security	Data protection service	Strategy (data leakage prevention, data loss prevention), subject (data in use, data in transit, data at rest), regularity

5.3. EA concerns and elements for category C

Fulfilling the rights of the data subject requires a thorough understanding of which external data sources are tapped, which data objects belong to which natural person and where these are stored, so that they can immediately be corrected, deleted or their processing be restricted on demand. Moreover, their electronic transmission to other organizations may be requested. Exemplary stakeholder concerns for this category are:

- Which external data sources are tapped to obtain additional personal data [7]? How reliable, secure and privacy-friendly are these data sources?
- How heterogeneous are our data objects and stacks formatted? How can we transmit personal data in a common format [19]?
- What are the implications for our business when limiting the processing of certain personal data?

To answer these concerns, we recommend modeling external data sources [2] as well as attributes relating to the format and exchange of data [4] as shown in Table 3:

Table 3. EA elements for category C

EA layer	EA artifact	Attributes
Business	Business process	Level of dependence on personal data
	Business service	Level of dependence on personal data
Data	Data object	Format, degree of de-identification
	Data stack	Format, degree of de-identification
	Database	Portion of personal data
	Data stream	Data exchange language (XML, JSON, REBOL)
	External data source	Type (social media, public web, sensor, machine log), reputation, availability, security certificate

5.4. EA concerns and elements for category D

Ensuring an adequate level of security and privacy along all layers of EA in context of privacy by design requires a continuous balance of costs and risks. Thus, a comprehensive understanding of actually implemented security measures is necessary, raising several concerns:

- What does our security concept look like [20]? Which security measures did we implement to prevent data thefts and unauthorized access [7, 8]?
- How do we ensure de-identification of personal data? Which measures do we use to encrypt and anonymize personal data [7]?
- Which application and infrastructure components require particularly high protective measures against cybercrime? How robust is our IT infrastructure to breakdowns and disruptive events [20]?

On the security layer, the de-identification method provides the needed algorithms for encoding data (see Table 4). While the authorization and authentication services are responsible for managing access to several EA elements, the infrastructure protection service has to continuously monitor and secure the technology layer, which requires a high transparency about the consistency of its embedded components [9, 15]. In addition, it is inevitable to understand where infrastructure elements are located and which organizational unit is responsible [14]. Preventive organizational and technical measures also require awareness of the composition of applications in order to identify security gaps untimely [2, 23].

Table 4. EA elements for category D

EA layer	EA artifact	Attributes
Business	Organizational unit	Description, number of actors
	Location	Description, country, region
Application	Application component	Class (module, procedure, GUI), lifecycle status (proposed, in development, live, phasing out, retired)
	Application function	Frequency, lines of code, level of automation
Technology	Infrastructure element	Level of virtualization, physical integrity, elasticity, scalability
	Hardware component	Type (CPU, main memory, hard disk, expansion card, drive, power supply unit), resilience, capacity, maturity
	Network component	Type (bridge, repeater, hub, cable), security standard (WEP, WPA, WPA2), security protocol (TKIP, CCMP), resilience, transmission rate, maturity
	System software	Class (operating system, utility software, middleware), version
Security	Security goal	Priority, success criteria
	Authorization service	Function (policy enforcement, policy distribution, policy control, role management), policy language (XACML), access control paradigm (ABAC, RBAC)
	Authentication service	Authentication method (graphical authentication, SAPA, SSO, port-knocking), standard (SAML)
	De-identification method	Type (pseudonymization, anonymization, suppression, generalization, encryption), cryptographic hash function (SHA-384, SHA-512), provided encryption scheme (AES, RSA, DSA, ECC)
	Infrastructure protection service	Function (intrusion detection, firewall, content filter, threat modeling, vulnerability analysis), regularity

6. Demonstration of the privacy-driven EA meta-model

EA meta-models shall provide “a common language and a clear view on the structure of and dependencies between relevant parts of the organization” [12]. They function as a template for EA models that are capable of answering specific concerns by prescribing permissible layers, artifacts, attributes and relations [21]. In this

way, EA meta-models enforce coherence and semantic rigor along derived EA models, which are preconditions for successful communication and documentation [12].

To develop our privacy-driven EA meta-model (see Figure 3), we interlinked the previously derived EA artifacts based on their logical interrelation as described in [21]. Additionally, we added elements capturing involved actors for highlighting the dependence and needed awareness of external parties when processing personal data. The security layer is arranged in parallel to the other layers, since it is responsible for appropriate protection mechanisms throughout the whole EA. The

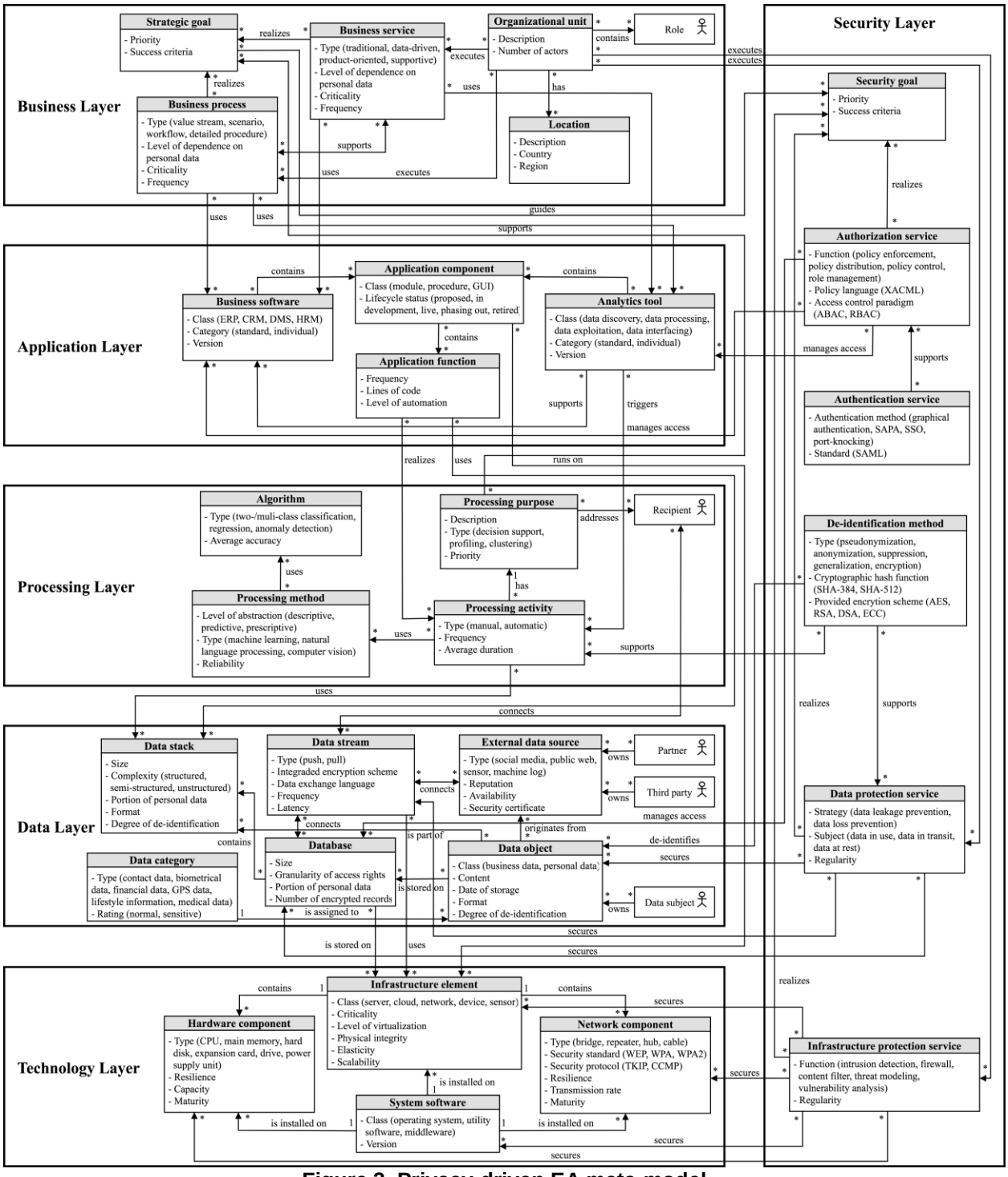

Figure 3. Privacy-driven EA meta-model

processing layer, providing the required transparency about each processing activity and the implemented logic of analytics according to the GDPR, is triggered by the application layer and relies on input from the data layer.

For demonstrating the meta-model, we exemplarily choose the first concern of each of the four categories.

Category A: *Why do we collect certain personal data? Which goals and business objects are affected by processing personal data and how important are these for the overall business?* The processing layer facilitates a complete documentation of all processing purposes, which is required by the GDPR. A processing purpose, for instance, might be the optimization of marketing in context of profiling or the acceleration of an application procedure. By linking the processing layer with the application layer and business layer, it becomes visible which strategic goals, business processes and services are supported by which processing purpose and based on their criticality and priority attributes, how important these are for the overall business.

Category B: *Which applications process, analyze, visualize and use personal data? Which methods and algorithms realize the implemented logic of analytics?* By modeling analytics tools in detail and linking them to supported business software and the processing layer, transparency about the implemented logical sequence of analytical processing can be achieved. Moreover, a well-documented processing layer clarifies which analytics abilities are actually available at all and which algorithms are responsible for processing which personal data.

Category C: *Which external data sources are tapped to obtain additional personal data? How reliable, secure and privacy-friendly are these data sources?* Capturing external data sources within the data layer and linking them to other EA elements provides information about their significance for the whole business. In addition, attaching privacy-related attributes to these data sources clarifies whether they are even suitable for integration and to which extent data streams need to be protected.

Category D: *What does our security concept look like? Which security measures did we implement to prevent data thefts and unauthorized access?* Detailing embedded services for authorization, authentication as well as data and infrastructure protection on the security layer and connecting them with the other layers raises awareness of the level of security within an EA. By visualizing the flow of personal data and modeling privacy-relevant attributes along the EA, it additionally becomes clear which EA elements require particularly stringent security measures.

Although we developed our meta-model based on privacy-related concerns, we argue that it is also capable of answering concerns related to the optimization and homogenization of EA, since it provides transparency about data-driven correlations and potentials.

7. Discussion

The increasing requirements on security and privacy, stemming from innovative technologies and modern regulations, such as the GDPR, pose complex challenges for the EAM. Realizing the continuous transformation of an enterprise necessitates an increasing focus on the identification of security gaps and greater consideration of privacy-related issues. Planning roadmaps of changes for EA evolution requires constant attention to potential impact on the protection of data [8] and an appropriate balance of risks and costs, since non-compliance with regulatory requirements can result in severe penalties, damage to the public image and far-reaching economic losses. As a result, security- and privacy-related aspects need to be reflected more closely in EA frameworks, EA patterns and EA meta-models, since these embody essential instruments of the EAM [19].

Additionally, the role of the enterprise architect is essential for compliance. To realize enterprise-wide data protection, enterprise architects are required to work closely with data protection officers. Given the diversity of privacy-related concerns as demonstrated, enterprise architects are particularly well suited to support the data protection officer's efforts due to their unique and fully integrated vantage point of an enterprise [7].

By providing our privacy-driven EA meta-model as a template for deriving current and target EA models, we aim to support enterprise architects and the EAM in performing the transformation and maintenance of the EA in a privacy-friendly manner and thus in ensuring continual compliance with regulations like the GDPR.

8. Conclusion

In their paper "15 Years of Enterprise Architecting at HICSS: Revisiting the Critical Problems", Kaisler and Armour state that "no papers addressed the co-development of a security architecture as an essential element of the EA" [14]. Moreover, they state that "additional artifacts are required in an EA: identification of security and privacy vulnerabilities, defensive technologies, and mitigating practices to ensure security and privacy compliance with appropriate regulations" [14]. In our paper, we referred to the GDPR as a highly topical regulation on data protection and developed a privacy-driven EA meta-model in order to address this research gap.

The results of this paper contribute to science and practice alike. From an academic perspective, they provide implications for additional research on the interplay between EA, security and privacy. In addition, they demonstrate a concern-driven approach towards transforming regulatory requirements into EA elements.

Moreover, the privacy-driven EA meta-model aims to bridge the gap between EPA, ESA and EA. For practice, the results highlight the usefulness of EA models for achieving compliance with the GDPR. The meta-model in particular provides guidance to gain transparency about the processing of personal data by proposing EA elements related to analytics, privacy and security. The identified four categories should also give enterprises an overview of their GDPR-related obligations.

The results of this paper are not without limitations. First, the EA concerns and elements were deductively derived by a content analysis of the GDPR considering additional literature. Performing case studies or expert interviews would provide further insights and lead to additional concerns. Second, future court decisions on GDPR-related issues might result in additional concerns and requirements on EA not considered in this paper.

Additional research is required on the integration of EA, security and privacy. Our future work will focus on refining the meta-model by transferring it to different application domains and on studying how these domains use EA modeling to achieve compliance with the GDPR.

9. References

[1] T. Huth, A. Faber, and F. Matthes, "Towards an Understanding of Stakeholders and Dependencies in the EU GDPR", *Proceedings of the MKWI 2018*, Lüneburg, Germany, pp. 338-344.

[2] P. Pääkkönen and D. Pakkala, "Reference Architecture and Classification of Technologies, Products and Services for Big Data Systems", *Big Data Research*, 2(4), 2015, pp. 166-186.

[3] O. Tene and J. Polonetsky, "Big Data for All: Privacy and User Control in the Age of Analytics", *Northwestern Journal of Technology and Intellectual Property*, 11(5), 2013, pp. 239-273.

[4] L. Xu, C. Jiang, J. Wang, J. Yuan, and Y. Ren, "Information Security in Big Data: Privacy and Data Mining", *IEEE Access*, 2, 2014, pp. 1149-1176.

[5] A. F. Westin, "Privacy and Freedom", Ig Publishing, New York, 2015.

[6] European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council", *Official Journal of the European Union*, 111, 2016, pp. 1-88.

[7] J. Bloomberg, "Enterprise Architects: Critical Resource for GDPR Compliance", 2017, available at: https://samu.io/enterprise-architects-critical-resource-gdpr-compliance/, accessed 26 May 2018.

[8] A. Campbell, "GDPR - Gaining Accountability using EA", 2017, available at: https://www.enterprisearchitects.eu/blog/-/blogs/gdpr, accessed 10 May 2018.

[9] The Open Group, TOGAF Standard, version 9.2, 2018.

[10] R. Fischer, S. Aier, and R. Winter, "A Federated Approach to Enterprise Architecture Model Maintenance", *Enterprise Modelling and Information Systems Architectures*, 2(2), 2007, pp. 14-22.

[11] D. Simon, K. Fischbach, and D. Schoder, "An Exploration of Enterprise Architecture Research", *Communications of the Association for Information Systems*, 32(1), 2013, pp. 1-72.

[12] J. Saat, U. Franke, R. Lagerström, and M. Ekstedt, "Enterprise Architecture Meta Models for IT/Business Alignment Situations", *Proceedings of the 14th IEEE EDOC*, Vitoria, Brazil, 2010, pp. 14-23.

[13] The Open Group, "TOGAF and SABSA Integration - How SABSA and TOGAF complement each other to create better architectures", 2011.

[14] S. H. Kaisler and F. Armour, "15 Years of Enterprise Architecting at HICSS: Revisiting the Critical Problems", *Proceedings of the 50th Hawaii International Conference on System Sciences*, Waikoloa, Hawaii, 2017, pp. 4807-4816.

[15] G. Karjoth, M. Schunter, and M. Waidner, "Privacy-enabled services for enterprises", *Proceedings of the 13th International Workshop on Database and Expert Systems Applications*, Aix-en-Provence, France, 2002, pp. 483-487.

[16] S. M. Oda, H. Fu, and Y. Zhu, "Enterprise Information Security Architecture: A Review of Frameworks, Methodology, and Case Studies", *Proceedings of the 2nd IEEE ICCSIT*, Beijing, China, 2009, pp. 333-337.

[17] M. Shariati, F. Bahmani, and F. Shams, "Enterprise information security, a review of architectures and frameworks from interoperability perspective", *Procedia Computer Science*, 3, 2011, pp. 537-543.

[18] Open Security Architecture, official website, available at: www.opensecurityarchitecture.org, accessed 16 May 2018.

[19] M. Lankhorst, "8 Steps Enterprise Architects can take to deal with GDPR", 2017, available at: https://bizzdesign.com/8-steps-enterprise-architects-can-take-to-deal-with-gdpr/, accessed 20 May 2018.

[20] L. Moné, "Mastering the GDPR with Enterprise Architecture", LeanIX white paper, version 1.1, 2018.

[21] L. Uzzle, "Using Metamodels to Improve Enterprise Architecture", *Journal of Enterprise Architecture*, 5(1), 2009, pp. 49-61.

[22] P. Mayring, "Qualitative Content Analysis", *Forum: Qualitative Social Research*, 1(2), 2000, pp. 1-10.

[23] Y. Mengke, Z. Xiaoguang, Z. Jianqiu and X. Jianjian, "Challenges and Solutions of Information Security Issues in the Age of Big Data", *China Communications*, 13(3), 2016, pp. 193-202.