Using Enterprise Architecture Models for Creating the Record of Processing Activities (Art. 30 GDPR)

Dominik Huth, Ahmet Tanakol, Florian Matthes

Department of Informatics, Chair of Software Engineering for Business Information Systems
Technical University of Munich
Garching, Germany
{dominik.huth, ahmet.tanakol, matthes} @tum.de

Abstract—The record of processing activities (RPA) is a central document in demonstrating compliance with the General Data Protection Regulation (GDPR). Article 30 of the GDPR specifies the information that has to be made available to the supervisory authority upon request. Currently, data protection management experts conduct their own data collection and maintain isolated RPAs. We show how existing Enterprise Architecture models can be augmented with the necessary information to maintain and generate an RPA. We evaluate the completeness and usefulness of the approach together with data protection management experts.

Index Terms—Enterprise Architecture Management, GDPR, data protection, record of processing activities, RPA, ArchiMate

I. Introduction

Information technology shapes the historical period that we live in the 21st century. The Information Age created a world where high-tech companies deliver services and products that are based on vast amounts of data, often from individuals. Leveraging and analyzing personal data have become major growth drivers, e.g. for targeted advertisements or personalized services. To ensure that this processing is carried out in accordance with the fundamental rights and freedoms of individuals, the European Union (EU) passed the General Data Protection Regulation (GDPR) in 2016 [1]. Just as privacy is a *set of protections against a related set of problems* [2], the GDPR can be seen as a legal toolbox for addressing the various facets of privacy.

Establishing and maintaining GDPR compliance requires a complete overview of the organization, an understanding of processes, applications and data flows, a vocabulary and model to abstract these concepts, and a method to obtain consistent and reproducible results. Since these are also objectives of Enterprise Architecture (EA) Management, we advocate the collaboration between EA management and the discipline that we call *data protection management (DPM)*.

The accountability principle is an essential part of the GDPR. It enforces organizations to ensure compliance and

This work has been sponsored by the German Federal Ministry of Education and Research (BMBF) grant 01IS17049 / UMEDA. The responsibility for the content of this publication lies with the authors.

demonstrate it to the supervisory authorities. One provision addressing the accountability principle is the obligation to maintain a record of processing activities (RPA) in accordance with Art. 30 GDPR [3]. The RPA has to be made available to the supervisory authority and serves to demonstrate GDPR compliance at company level [1]. Given certain limitations, such as employee count, every company has to keep an RPA in writing or in electronic form. According to Art. 30, it must contain the following information:

- The name and contact details of the controller
- The name of the data processing activity
- The purposes and lawful basis of the processing activity
- The categories of data subjects and personal data
- The categories, names and contact details of recipients to whom the personal data have been or will be shared (both internal and external)
- The identification of third countries or international organization in the case of transfers of personal data
- Retention period of different categories of data
- A description of the technical and organizational security measures

As confirmed in interviews with data protection management experts, the creation of an RPA is still a highly manual process. The responsible person, e.g. the data protection officer (DPO), gets in touch with stakeholders within the organization and collects the required information in personal interviews or by sending out questionnaires, often via E-Mail. The collected data is then consolidated in spreadsheets and updated upon request. This usually takes place on a yearly basis or when triggered by special events, such as the May 25th deadline of the GDPR. The current approach of RPA creation is shown in figure I.

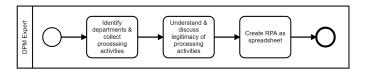


Fig. 1. State of the art process for creating an RPA.

EA modeling describes different aspects of an organization by using a common vocabulary among stakeholders. This work focuses on EA support for the creation of the RPA. We envision the collection of information that is relevant for the RPA in a consolidated EA model. Since EA models are maintained for other reasons than DPM, we propose to enrich these models with data protection information. The consolidated data in enhanced EA models can then be used for the automated creation of the RPA, which facilitates the work of data protection management experts. At the same time, a larger number of stakeholders promotes a higher data quality in the EA model and thus creates a benefit for EA management as well. Thus, the research goal of this work is to **support the automated creation of an RPA with EA**. We seek to cover the following research questions:

- RQ1: How can EA models support the automated creation of an RPA?
- **RQ2**: What are the advantages and challenges of the process we propose?

The remainder of this work is structured as follows: We present the necessary foundations and an example EA model in section II and explain our research method in section III. Section IV provides a detailed model of the requirements for an RPA and the information properties for each model element. In section V, we present an approach to leverage the model for the automated creation of an RPA and discuss the feedback from data protection management experts. After a short presentation of related research in section VI we conclude in section VII.

II. BACKGROUND AND MOTIVATING EXAMPLE

A. Enterprise Architecture Management

Enterprise Architecture, according to [4], is a coherent whole of principles, methods, and models that are used in the design and realization of an enterprises organizational structure, business processes, information systems, and infrastructure. EA management is a continuous and iterative process to maintain and improve the alignment of business and IT [5]. Specific information has to be presented to the respective stakeholders in accessible form [6]. In this work, we consider stakeholders from the DPM discipline, and thus propose a modeling approach that can be leveraged for the concerns of DPM experts. Models, which [7] defines as purposefully created artifacts, can be used to achieve this objective. According to [8], an EA process is an iterative process with the steps model & collect (1), communicate, explain, involve (2) and adapt & reflect (3).

B. ArchiMate

ArchiMate is a graphical language for representing Enterprise Architectures that was developed in a co-operation of partners from government, industry and academia and was later transferred to The Open Group [9]. The core framework divides Enterprise Architecture into three layers to describe it. Each layer is represented by a color as follows; yellow color represents elements from the *Business Layer*, blue color represents elements from the *Application Layer* and green color represents elements from the *Technology Layer*. Each layer comprises various modeling elements that can be classified into three different aspects, namely *passive structure*, *behavior* and *active structure*. For each element, a symbol is used to describe the type of the element, and connections between different elements are achieved through the relationships based on the types of the elements. Additional elements of the language that are provided by the full framework include concepts for describing the motivational elements such as requirements and goals for an architecture. Reasons behind architectural decisions can be expressed by using the motivational layer, and the elements that belong to the motivational layer are represented by purple color.

C. Example Enterprise Architecture Model

In this example, layers, aspects, and their interactions will be covered based on the previous explanations. Order Placement is a business process that is composed of sub-processes. There are triggering relationships between these business processes (e.g. Enter Delivery Information triggers the Create Order process, which in turn triggers Take Payment). The Order Placement business process realizes the Give Order business service. Customer, modeled as a business actor, is assigned to a business role called Application User. Give Order serves to Application User. Customer is associated with personal information. Customer Information, Order Information, and Credit Card Information are realized by the data objects from the application layer. Customer Data, Order Data and Credit Card Data are accessed by the application components. Order Application, Financial Application, and Notification Application are associated with the application functions, namely *Order* Creation Function, Payment Function, Notification Function. Each of the application functions is realized by application services which serve different business processes in the Order *Placement* process. The described EA model is reflected in the yellow and blue elements in Figure 2. The purple elements of Figure 2 will be explained in section IV.

III. RESEARCH PROCESS

Our research process was composed of four steps. First, we identified the necessary information requirements for an RPA from literature and practical guidance documents and consolidated these requirements into the list that is given in section IV. The requirements were then discussed and consolidated among two researchers.

In the second step, we modeled these requirements as goal model in ArchiMate. We created a simple EA model in ArchiMate, which is comprised of only one main business process and its supporting applications. The model elements were linked to the respective requirement elements and enhanced with the necessary information properties to fulfill these requirements.

The third step was the design of an approach to leverage the information in the ArchiMate model. We transferred the ArchiMate model to a graph database to facilitate querying the elements. It is possible to retrieve all necessary information for an RPA with one relatively simple query. The query result could be consumed by an application that serves as an RPA front-end.

We then communicated our model and the approach to data protection management experts. In the subsequent interviews and discussions, we collected feedback and suggestions.

IV. MODELING RPA REQUIREMENTS

A. Identification of Requirements

We analyzed Art. 30 and Recital 82 of the GDPR [1], as well as various RPA templates by national data protection authorities (e.g. the template by the UK ICO¹), in order to identify the information requirements for the RPA. This is the information that has to be represented in an EA model in order to automatically generate an RPA. The analysis resulted in the properties that are displayed in the respective column in table I.

B. Goal model for RPA

In the terminology of the ArchiMate motivation extension, the GDPR can be considered as an *external driver*. A *goal* that is associated with this driver is the documentation of GDPR compliance. This goal is fulfilled by the *outcome*: *documents to show GDPR compliance*, which in turn has the *requirement*: *Create Record of Processing Activities*. We show a detailed decomposition of this requirement in Figure 2. Since the processing activity is at the core of the RPA, we focus on describing and modeling the processing activity. The information requirements regarding the roles of data controller, data processor or data protection officer can be handled in a similar way.

C. Information Requirements

In order to be able to extract the necessary information for these requirements, we extend the sample model in section II with actors and roles. An additional element is the *Organization*, which is composed of other business actors, namely *Person X* and *Departments*. For the sake of simplicity, *Organization* acts as both the data controller and data processor in our example. *Person X* is designated as *Data Protection Officer*. Information about the data protection officer is stored in the *Person X Information* business object which is realized by *Person X Data*. For the rest of the business actors, their information is represented in the same way.

While some basic information (e.g. the role name) can be extracted from the element names, we represent the necessary information by adding properties to the model elements. We enrich the model by defining key-value pairs for the elements and relationships.

As the basis for the processing activity, we use the model element *business process*. The name of the business process can be used directly as the name of the processing activity.

¹https://ico.org.uk/media/for-organisations/documents/2172937/gdpr-documentation-controller-template.xlsx, accessed 05/16/2019

Further, a business process adds context to data objects, because it processes them in a certain manner. A business process usually has a purpose, which can either be derived from the name itself (as proposed by [10]), or added to the business process as another information property. Instead of additional model elements, we add a textual description and the legal basis for processing as properties to the business process.

A business process or function can be composed of other processes and functions. We consider the overarching processing activity (*Order Placement*) to have the adequate granularity for the RPA.

In addition to the categories of internal recipients or external recipients, the department, the role, the name and other information of the responsible person can be extracted from the model elements that have direct or indirect relationships with the business role element. The responsible person of the processing activity and the category of the internal recipients are found by following the relationships represented in Figure 2.

The *Order Placement* process uses three types of data: *Customer Information*, *Credit Card*, *Order Information*. These business objects are consumed and produced by *Order Placement*. Information about the categories of personal data, the collected data and the retention periods can be defined in these business objects.

Table I assigns all identified information properties to Archimate model elements.

V. USING THE MODEL FOR THE AUTOMATED CREATION OF AN RPA

During the evaluation interviews with the DPM experts, we also discussed the current approach for the creation of an RPA. It currently involves a large effort for data collection and communication of behalf of the DPM expert. Processing activities and the required information are obtained in direct interviews with the departments and entered into spreadsheets or specialized solutions for maintenance of an RPA. This record is updated on a regular basis (e.g. every year) or before special events, such as audits.

In parallel, Enterprise Architects collect information from the same departments, but with other objectives. We propose an approach to combine these data collection efforts and consolidate the necessary information for both EA management and data protection management in a unified model, such as the one we proposed in the previous section. The EA model then serves as the primary source of information for the RPA and provides additional value to both disciplines. This approach is visualized in Figure V.

A. Automated Creation of RPA

We use the Archimate model that we presented in section IV and describe a step-by-step process how to automatically create the entry for our example process in the RPA. We show how a single query can return all the necessary information properties that we derived in section IV. The result can be

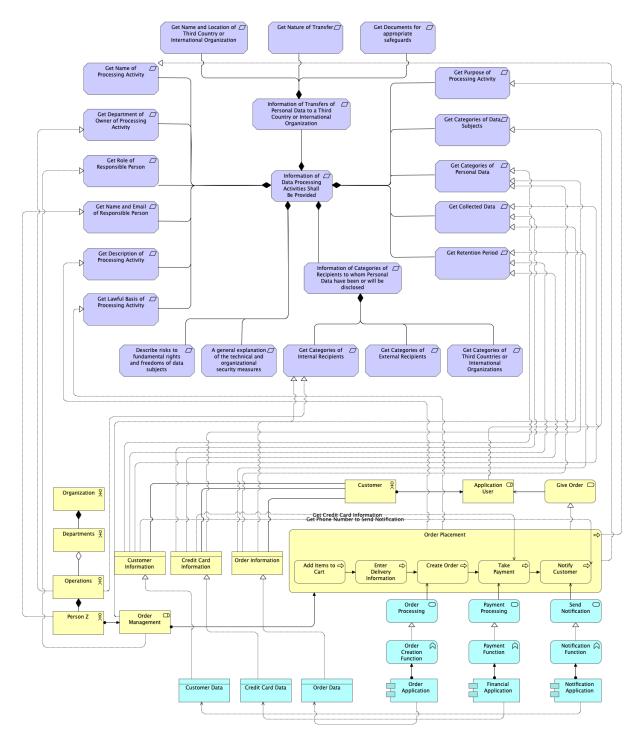


Fig. 2. Goal model of the RPA with realization relationships to our example EA model

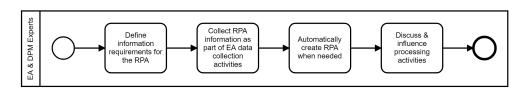


Fig. 3. Proposed process with EA models as primary source of information.

ArchiMate Element	Property
Business Actor	Department of Owner of Processing Activity, Name of Responsible Person, Email of Respon- sible Person, Category of Data Subject, Cate- gory of Internal Recipient, Category of External Recipient, Category of Third Country or Inter- national Organization, Name of Third Country or International Organization, Location of Third Country or International Organization
Business Object	Name of Data Protection Officer, Name of Responsible Person, Email of Responsible Person, Category of Data Subject, Category of Personal Data, Collected Data, Retention Period, Category of Internal Recipient, Category of External Recipient, Category of Third Country or International Organization, Name of Third Country or International Organization, Location of Third Country or International Organization
Business Role	Role of Responsible Person, Category of Data Subject, Category of Internal Recipient, Cate- gory of External Recipient, Category of Third Country or International Organization
Location	Location of Third Country or International Organization
Business Process	Name of Data Processing Activity, Description of Processing Activity, Lawful Basis of Processing Activity, Purpose of Processing Activity
Contract	Documents for appropriate safeguards
Representation	Nature of Transfer

TABLE I

OVERVIEW OF ARCHIMATE ELEMENTS AND PROPERTIES FOR THE REALIZATION OF THE REQUIREMENTS OF THE DATA PROCESSING ACTIVITIES

used in a straightforward way by to create an RPA, either in spreadsheets or in an application that displays the queried data.

The Archimate model we presented includes many relationships between the individual elements. Since graph databases handle relationships as first class citizens, we decided to use the graph database Neo4j² for our demonstration. Various practitioners have proposed transferring ArchiMate models to Neo4j and using its advanced querying functionalities to explore the models. We follow the approach described by [11], but refer to the source for details. This results in a graph representation of the EA model.

Instead of displaying all the graph elements and relationships, we can create a visualization solely for understanding how the necessary data can be retrieved to generate a record of processing activities. To analyze which elements realize the requirements that we modeled in ArchiMate, we can use the following query, whose results are shown in Figure 4:

By improving the previous query, we can remove some of the irrelevant properties. The following query creates a map of the properties and removes the keys that are given as the second parameter for the *removeKeys* function.

The result contains all the necessary information for the processing activity *Order Placement* in the RPA. If multiple business processes are represented in the EA model as we specified it, the query extracts the complete record of processing activities of an organization. This information can be consumed by an application that ensures an up-to-date version of the record of processing activities at any time.

B. Evaluation with DPM experts

We conducted four semi-structured interviews with data protection experts and held a focus group interview with another four data protection experts at a local data protection interest group. The eight interviewees all work in data protection management in smaller or larger organizations. We did not obtain consent to use any kind of identifiable information about the experts and the affiliated organizations. Thus, we consider this section as an informal indicator for the usefulness of the model and the associated approach.

In both the interviews and the focus group interview, we first presented the core concepts of EAM. Some experts were familiar with enterprise models on an abstract level, but not with detailed technical models, such as ArchiMate. We presented the ArchiMate model of the *Order placement* example and showed how the information can be extracted from the model. Subsequently, we entered into an open discussion about the feasibility of the approach in RPA creation. When necessary, we asked stimulating questions, such as:

- Does the model capture all the necessary information for an RPA?
- What are the advantages or challenges of this approach?
- What should be the functionalities of a tool that automatically creates an RPA?

As a general observation, some of the data protection officers we interviewed did not know if their organizations had an EA department. Our presentation encouraged them to actively investigate the existence of EAM within their organizations, because they recognized the value in the relationship knowledge that is represented in EA models.

Regarding the **model**, the experts did not analyze the RPA completeness on a per-attribute basis, but pointed out

²https://neo4j.com/

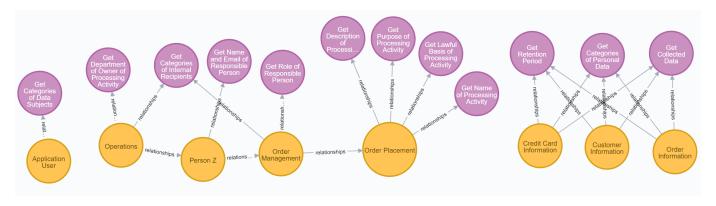


Fig. 4. Graph representation of the requirements and the realizing EA elements.

the parallels between the EA model elements and the RPA information requirements. EA processes hint at processing activities, and applications are an entry point to find out which processing activities they support. Since not all data processing encompasses processing of personal data, it was encouraged to model a distinction of processes or applications, depending on whether or not they process personal data. Overall, the experts did not consider it as feasible to have all the required information stored in one model, but mentioned the usefulness in having subsets of the necessary attributes in the model.

A challenge that organizations face is that they want to know which data can be used for which purpose. Legal departments try to avoid data usage as much as possible to reduce compliance risks. On the other hand, many business opportunities are based on leveraging personal data. Therefore, our interviewees suggested enriching the data objects with the allowed processing purposes.

The **approach** of using EA models as source of information for the RPA was evaluated as extremely helpful. Any company-wide account of applications or processes can provide a good starting point for DPM activities. Automation of this process was seen as relative, because the data collection still has to take place manually. Our interviewees pointed out that creating a record of processing activity is not only about data. A data protection officer has to decide whether a data processing activity is legitimate or not, and the limited amount of information that is *reported* in an RPA is in no way sufficient for this decision. Modeling this information is not an option, since the decision highly depends on the processing activity and is made on a case by case basis. Rather, the DPM experts would use the contact information to discuss the activity directly with the responsible person if doubts arise.

Critical feedback for this approach was that it is too technical. The DPM experts understood the general purpose of EA and the information that our example model represents, but found the Neo4j tool and the representation as a flat table hard to understand.

Additional **remarks and suggestions** included the combination of data protection tools with EA management tools, such that whenever an expert from the EA or DPM disciplines updates information, the information is updated for all

stakeholders for this information.

VI. RELATED WORK

We are only aware of a small number of publications addressing EA and GDPR compliance in general and the record of processing activities specifically. We present these publications in decreasing order of relatedness. Additionally, we present work that does not focus on the GDPR, but addresses EA and regulatory compliance.

In [12], the authors present a systematic approach for GDPR implementation based on EA. The first element of the approach is the creation of an RPA. In contrast to our work, the authors define phases for a project plan and describe the roles and specific tasks that each stakeholder has to fulfill. This is specified without the help of a modeling language. The method uses the list of applications from the EA repository as a starting point, but suggests manual reorganization into spreadsheets.

A discussion of the relationship between EA and the GDPR can be found in [13]. The authors argue that basic EA elements, such as actors, data or processes are essentially also GDPR elements, but they do not provide details of how EA can support in a specific GDPR task. In [14], the authors derive major obligations and stakeholder concerns and propose a privacy-driven metamodel that incorporates the information for the identified aspects and thus promotes privacy compliance. Other work with respect to GDPR requirements from an IT management perspective focuses on the development of GDPR compliant software [15], data protection impact assessments (DPIA) [16], [17], or data portability [18],

[5] maps security and risk management concepts to the ArchiMate modeling language. A similar approach can be found in [19], where a critical discussion and a possible redesign of ArchiMate's risk and security overlay (RSO) is proposed.

From the EA industry perspective, BiZZdesign proposes an approach for RPA creation that also involves extending the metamodel of the application [20]. However, it is not clear to which model elements these information properties are attached.

VII. DISCUSSION & OUTLOOK

We have shown that existing EA models can be enhanced with all relevant information that is necessary for creating an RPA for *reporting* purposes. The concepts of processing activities and roles align well with EA elements, and existing modeling languages, such as ArchiMate, can be used for capturing information that is relevant for data protection management. An EA model in its machine-readable format can be used for storing this information. From the model, an up-to-date RPA can be generated at any time. We showed this by defining an approach to query all RPA information in a structured format. The data in this structured form can be displayed in a custom-built application or exported to a spreadsheet.

Our approach was discussed with several DPM experts in direct interviews and a focus group interview. The feedback indicates that the EA models themselves provide value to DPM, because they capture relationship knowledge of the organization that would otherwise have to be collected by DPM experts with additional effort. Most importantly, the employed applications provide an entry point to identify potential processing activities and collect more information. DPM experts were convinced of the positive contribution that EA can make to DPM. They suggested combined tools for EAM and DPM with specialized views for data protection, such that both disciplines profit from more recent models. However, the approach is rather an improvement than a full solution for RPAs, because DPM experts need more information to reason about processing activities. Further, the information properties are based on literature rather than court decisions, so the responsibility still lies with the DPM experts.

From our findings, we conclude that EA should be promoted as an enabler for privacy compliance. Both EAM and DPM can benefit from combined data collection efforts, and making EA information available to DPM experts is a first step.

Future work includes understanding how EA experts already support data protection management. We are currently conducting interviews with EA experts that are focused on the tasks that EA supported with and the collaboration with data protection management during the preparation period for the GDPR. First results indicate that industry leveraged EA documentation to a varying degree in the creation of the RPA, from one-time exports of application lists to inclusion of the DPM experts in the EA documentation process. Our goal is to further advance this collaboration by identifying successful practices, and establishing and evaluating guidelines for efficient data protection management.

REFERENCES

- [1] European Union, "Regulation 2016/679 of the European parliament and the Council of the European Union," pp. 1–88, 2016.
- [2] D. J. Solove, "I've got nothing to hide and other misunderstandings of privacy," San Diego I. Rev., vol. 44, p. 745, 2007.
- [3] C. Tikkinen-Piri, A. Rohunen, and J. Markkula, "EU General Data Protection Regulation: Changes and implications for personal data collecting companies," *Computer Law and Security Review*, vol. 1, no. 2017, 2017.

- [4] M. Lankhorst, Enterprise Architecture at Work. Heidelberg: Springer, 2005.
- [5] E. Grandry, C. Feltus, and E. Dubois, "Conceptual Integration of enterprise architecture management and security risk management," *IEEE International Enterprise Distributed Object Computing Workshop*, EDOC, pp. 114–123, 2013.
- [6] H. Jonkers, M. M. Lankhorst, H. W. L. ter Doest, F. Arbab, H. Bosma, and R. J. Wieringa, "Enterprise architecture: Management tool and blueprint for the organisation," *Information Systems Frontiers*, vol. 8, no. 2, pp. 63–66, 2006.
- [7] H. A. Proper, M. Bjeković, B. Van Gils, and S. De Kinderen, "Enterprise Architecture Modelling: Purpose, Requirements and Language," in IEEE International Enterprise Distributed Object Computing Workshop, EDOCW, 2018, pp. 162–169.
- [8] M. Hauder, S. Roth, C. Schulz, and F. Matthes, "Agile Enterprise Architecture ManagementAn Analysis on the Application of Agile Principles," in 4th International Symposium on Business Modeling and Software Design, 2014, pp. 38–46.
- [9] A. Josey, M. M. Lankhorst, I. Band, H. Jonkers, and D. Quartel, "An introduction to the ArchiMate 3.0 Specification," The Open Group, Tech. Rep. June, 2016.
- [10] D. Basin and T. Hildebrandt, "On Purpose and by Necessity: Compliance under the GDPR," Financial Cryptography and Data Security 2018, pp. 1–18, 2018.
- [11] L. Lazarevic, "Using Neo4j to explore your ArchiMate model," accessed: 2019-05-14. [Online]. Available: https://ljulazarevic.github.io/ArchiMateNeo4j1.html
- [12] H. Koç, K. Eckert, and D. Flaig, "Challenging the general data protection regulation (gdpr) with enterprise architecture management (eam)," *HMD Praxis der Wirtschaftsinformatik*, no. 5, pp. 942–963.
- [13] P. Rozehnal and V. Novak, "The Core of Enterprise Architecture as a Management Tool: GDPR Implementation Case Study," in 26th Interdisciplinary Information Management Talks. Kutná Hora, Czech Republic: Trauner Verlag, 2018, pp. 359–366.
- [14] F. Burmeister, P. Drews, and I. Schirmer, "A Privacy-driven Enterprise Architecture Meta-Model for Supporting Compliance with the General Data Protection Regulation," in 52nd Hawaii International Conference on System Sciences, 2019, pp. 6052–6061.
- [15] Y. S. Martin and A. Kung, "Methods and Tools for GDPR Compliance Through Privacy and Data Protection Engineering," in 3rd IEEE European Symposium on Security and Privacy Workshops, EURO S and PW 2018. IEEE, 2018, pp. 108–111.
- [16] F. Bieker, M. Friedewald, M. Hansen, H. Obersteller, and M. Rost, "A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation," in 4th Annual Privacy Forum 2016, no. LNCS. Cham: Springer, 2016, pp. 21–37.
- [17] R. Alnemr, E. Cayirci, L. Dalla Corte, A. Garaga, R. Leenes, R. Mhungu, S. Pearson, C. Reed, A. S. de Oliveira, and D. Stefanatou, "A data protection impact assessment methodology for cloud," in *Annual Privacy Forum*. Springer, 2015, pp. 60–92.
- [18] D. Huth, L. Stojko, and F. Matthes, "A Service Definition for Data Portability," in 21st International Conference on Enterprise Information Systems, Vol. 2, 2019, pp. 169–176.
- [19] T. Prince Sales, J. P. Andrade Almeida, S. Santini, F. Baiao, and G. Guizzardi, "Ontological analysis and redesign of risk modeling in archimate," 2018 IEEE 22nd International Enterprise Distributed Object Computing Conference, EDOC 2018, pp. 154–163, 2018.
- [20] J. Niehof, "Create your GDPR registers with BiZZdesign Enterprise Studio," 2017. [Online]. Available: https://bizzdesign.com/blog/createyour-gdpr-registers-with-bizzdesign-enterprise-studio/