DEPARTMENT OF INFORMATICS

TECHNICAL UNIVERSITY OF MUNICH

Master's Thesis in Information Systems

An Information Model as a Basis for Information Gathering to comply with Data Portability according to GDPR Art. 20

Laura Stojko

DEPARTMENT OF INFORMATICS

TECHNICAL UNIVERSITY OF MUNICH

Master's Thesis in Information Systems

An Information Model as a Basis for Information Gathering to comply with Data Portability according to GDPR Art. 20

Ein Informationsmodell als Basis für die Informationserhebung zur Datenportabilität nach Art. 20 DSGVO

Author:	Laura Stojko
Supervisor:	Prof. Dr. Florian Matthes
Advisor:	Dipl. Math.oec. Dominik Huth
Date:	November 15, 2018

I hereby declare that this thesis is entirely the result of my own work except where otherwise indicated. I have only used the resources given in the list of references.

Munich, 15. November 2018 Laura Stojko

Abstract

With the announcement of the General Data Protection Regulation (GDPR) by the European Union, data privacy laws shall be harmonized in European countries. In the digital era, personal data is of high significance for companies, especially within customer data-driven industries (e.g. social media platforms). Due to the new importance of personal data and its usage, the awareness for data privacy is increasing among people. Thus, good data privacy management is of high relevance for companies and customers. This thesis focuses on article 20 of the GDPR, which describes data portability as one essential aspect for new rights of the data subject. Data portability enables customers to receive their personal data from companies and transfer it to others. Thereby, a level playing field within the market is supported. To assure data portability in businesses, companies face the challenge to establish new processes and apply new techniques. An extremely high workload is expected as customers can request reports about their saved personal data. In order to fulfill the customers' demands, the relevant personal data has to be gathered from data sources. In particular, this data gathering process is highly complex and needs a structured procedure. Thus, this thesis develops an information model which can be used as a basis for collecting the required personal data. To develop the information model, typical approaches are discussed on the basis of literature. To ensure compliance with GDPR Art. 20, a requirement analysis for regulations is conducted. Further, the quality of the model is analyzed and improved by expert interviews, which are an essential part of evaluating the information model.

Contents

List of Figures

List of Tables

List of Abbreviations

API	Application Programming Interface
Art.	Article
B2B	Business to Business
CRUD	Create, Read, Update, Delete
CSV	Comma-Separated Values
DC	Data Controller
DCMES	Dublin Core Metadata Element Set
DS	Data Subject
DTP	Data Transfer Project
DWH	Data Warehouse
EAM	Enterprise Architecture Management
EDI	Electronic Data Interchange
EDIFACT	EDI For Administration, Commerce and Transport
ERP	Enterprise-Resource-Planning
EU	European Union
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
HTML	Hypertext Markup Language
ID	Identification
IRI	International Resource Identifier
IT	Information Technology
JSON	JavaScript Object Notation
MOF	Meta Object Facility
NISO	National Information Standards Organization

PDF	Portable Document Format
RDF	Resource Description Framework
RNLS	Restricted Natural Language Statement
SEPA	Single Euro Payments Area
SME	Small and Medium Enterprise
SQL	Structured Query Language
TOGAF	The Open Group Architecture Framework
U.S.	United States
UK	United Kingdom
UML	Unified Modelling Language
URI	Uniform Resource Identifier
W3C	Worl Wide Web Consortium
XML	Extensible Markup Language

Outline of the Thesis

Part I: Introduction and Theory

CHAPTER 1: INTRODUCTION

This chapter presents an overview of the thesis and its purpose. Furthermore, it introduces the motivation for the master's thesis topic and discusses current issues in data privacy laws. Also, a definition of important terms and an introduction to significant articles of the GDPR are given.

CHAPTER 2: DATA PORTABILITY (GDPR ART. 20)

This chapter explains a requirement analysis approach for compliance regulations and policies. Further, a requirement analysis by the previous presented approach is performed for data portability according to GDPR Art. 20. The article and how to extract the requirements are described. The result of this chapter is a list of necessary information to comply with GDPR Art. 20.

CHAPTER 3: INFORMATION MODEL BASICS

The content of this chapter discusses the basics of information models. In detail, basic terms of the information model and the information modelling process as well as the information model evaluation are introduced. Furthermore, it contains a discussion of two approaches for the basis of an information model, the document-based and business-object based one.

Part II: Information Model Development

CHAPTER 4: INFORMATION MODEL DEVELOPMENT

This chapter discusses the development of the information model. Therefore, the information model type, abstraction level, the information modelling process and the modelling language are discussed to find the best fit for the underlying problem domain. It concludes with the first draft of the information model.

CHAPTER 5: INFORMATION MODEL EVALUATION

This part describes how the first draft of the information model is evaluated and how the evaluation took place. Moreover, the expert interviews are summarized by reflecting the key points. In the end, the results from the evaluation are adapted to the information model, subsequently followed by the presentation of the final information model.

CHAPTER 6: INFORMATION MODEL ENHANCEMENTS

This chapter focuses on enhancement possibilities of the final information model. Therefore, several perspectives are discussed: The focus on an information model with detailed implementation details, the consideration of other articles of the GDPR and the actual transfer of data from one controller to another.

CHAPTER 7: CONCLUSION

The conclusion summarizes the thesis' outcomes and matches these to the previously defined research questions. An overview of the most important key elements of the thesis is given. Also the enhancement possibilities are emphasized and summarized to discuss the thesis' outcomes.

Part I.

Introduction and Theory

1. Introduction

On 27 April 2016, the European Parliament announced a new data privacy regulation called General Data Protection Regulation (GDPR) which became effective on 25 May 2018. The objective of this new regulation is to harmonize data privacy laws in European countries [16].

In the digital era customer data is used for several processing activities which might not agree with the customer's objectives. The customers' search history, for example, is analyzed to offer the best-fitting advertisement in comparison to their competitors [17]. From the customers perspective the question arises, how the company specialized their offers exactly for their needs. It makes customers feel monitored, controlled and may cause fear of losing control over their data. People's awareness of data privacy is increasing and influencing disciplines such as information systems, law, marketing, philosophy, psychology and economics [18]. Customers are more critical of companies and their data privacy management due to several data breaches in the past [19]. Privacy breaches occur regularly and represent risks to companies and their data handling [17]. Regardless, from the perspective of the company, data privacy stands for workload and costs. The advantage of good reputation and trust coming along with good data privacy management is often forgotten [17] [20], even though this can lead to higher customer satisfaction and most likely to an advantage over competitors. To point out the high importance of data privacy, the European Parliament discussed and finally published the GDPR in order to adapt to the digital age. Moreover, they are trying to improve data retention in Europe by encumbering companies who operate in the European Union with higher privacy requirements.

Along with the GDPR, the increased penalty [16] for data privacy breaches is a frequently discussed topic. Now, a severe privacy breach may cost a company the highest amount of either 20 millions or 4 percent of the company's worldwide total annual revenue. For a less serious breach the penalty may amount to either 10 millions or 2 percent of the company's worldwide total annual revenue. Again, the higher amount has to be paid by the company. The high penalties are the main reason for company's increasing awareness of GDPR compliance. As the GDPR is binding for companies operating in Europe or with data from European citizens, almost every company is affected by the regulation.

The GDPR covers lots of interesting topics. This thesis focuses on the high potential article of data portability (GDPR Art. 20) for improving data privacy. This article enables data subjects to get their data from a company and transfer it to another. The focus of this thesis is set on the data gathering procedure which discovers the data required for data portability compliance. The specific aspects are described in section 1.1.2. Data portability provides new data usage possibilities for data subjects, but also challenges companies to establish new processes and techniques for meeting the requirements. In summary, an extremely high workload is expected for complying with the GDPR. Data portability contributes to this high workload and costs as well.

1.1. Articles of the General Data Protection Regulation

Before focusing on some essential articles for this thesis's purpose, the structure of the GDPR is presented. In total, the GDPR groups the articles into the following topics [16]:

- General provisions (Art. 1 to 4)
- Principles (Art. 5 to 11)
- Rights of the data subject (Art. 12 to 23)
- Controller and processor (Art. 24 to 43)
- Security of personal data (Art. 32 to 43)
- Transfers of personal data to third countries or international organizations (Art. 44 to 50)
- Independent supervisory authorities (Art. 51 to 59)
- Cooperation and consistency (Art. 60 to 76)
- Remedies, liability and penalties (Art. 77 to 84)
- Provisions relating to specific processing situations (Art. 85 to 91)
- Delegated acts and implementing acts (Art. 92 and 93)
- Final provisions (Art. 94 to 99)

The structure shows the broad scope of the GDPR. The *rights of the data subjects* are most interesting for this thesis. These are of high interest to people, as they provide interesting possibilities on how to handle and control their data. In order to take advantage of the GDPR, information for enforcement is required. For fulfilling data portability requirements (GDPR Art. 20), data controllers need to know which data is of relevance for the transfer: the personal data of a data subject. Thus, the next subsections introduce the definition of personal data according to Art. 4 and data portability of the GDPR.

1.1.1. Article 4: Definition of personal data

This part introduces and explains the definition of personal data according to the GDPR Art. 4:

> *" "personal data" means any information relating to an identified or identifiable natural person ("data subject"); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person; " [16]*

Examples for personal data are names, addresses, date of birth, IP-addresses, bank account number. Using electronic services, personal data are email addresses, passwords and website data, like cookies. Also, location data, surfing behaviour and interests are personal data. All data categorized as personal data are affected of data portability (GDPR Art. 20). Thus, it is important to identify the correct information.

1.1.2. Article 20: Data portability

This section points out the importance of data portability according to Art. 20 of the GDPR. The article is categorized as a right of the data subject. The right enables data subjects to receive personal data which was provided by them to the data controller. Moreover, the provided format for the received personal data should be structured, commonly used and machine-readable. Further, the extract can be transmitted to another controller either from the data subject themselves or directly, if technically feasible. The main objectives of data portability are to empower and support the user, while giving them control over their data. By enabling easy data transfer, the companies have to stand out with other advantages than, for example, the price. Section 2.2 displays a deep discussion of the article's structure and requirements.

1.2. Research Outline

This section emphasizes the research goal, questions and method. Before enabling data transfer from one data controller to another, the company needs to identify the personal data that is actually in scope for the transmission. Thus, the research goal of this thesis is the development of an information model that supports the data gathering procedure to comply with GDPR Art. 20. Several research questions are necessary to achieve the research goal:

1. What are existing approaches of information models?

   A literature review identifies existing information model approaches based on either documents or business objects. Previously, the literature discusses the definition of an information model, the information modelling process and information model evaluation methods.

2. What are the requirements for the fulfillment of GDPR Art. 20?

   A literature review of a requirement analysis approach for compliance regulations helps to answer this research question. After identification of the approach, a requirement analysis of data portability (GDPR Art. 20) is conducted. As a result, the necessary information properties for data portability fulfillment are identified.

3. What is a possible information model covering the requirements of GDPR Art. 20?

   Combining the results from the first and second research question, the insights from the information model literature review and the necessary information for compliance with data portability result in a first information model draft.

4. How useful and applicable is the developed information model?

   The resulting information model draft is the basis for the conduction of expert interviews. Within the expert interviews, a discussion helps improving the usage and application of the information model. Thus, each interview outcome is iteratively improving the information model. Further, the outcome of the literature review regarding information model evaluation contributes to the interview content and structure.

In summary, the theoretical part of the thesis uses literature review as the essential research methodology to assure a good fundamental understanding of the key topics. The practical part uses findings of the information modelling literature review for the modelling of the information model. Also, the requirements for complying with GDPR Art. 20 are considered in the development of the information model. Furthermore, the information model evaluation focuses on expert interviews as methodology. Those identify information model enhancements, applications and improvements. Also, additional information about the general opinion regarding data portability and the GDPR are included in the interviews result.

2. Data Portability (GDPR Art. 20)

Data portability according to GDPR Art. 20 is part of a regulation which focuses on data subject's rights that require a deep analysis for good understanding. For this purpose, this section describes a method for the requirement analysis of such regulations and policies which is the procedure of Breaux [21]. Further, the method is applied to the analysis of data portability of the GDPR Art. 20. Moreover, the section's result includes the identified information properties as a result of the requirement analysis.

Only the knowledge about the requirements for fulfillment of GDPR Art. 20 is not sufficient to fully understand the complexity of the article's functionalities. Therefore, the challenges that come along with the regulation are discussed as well. As data portability in general has already been part of other directives, the last part of this section introduces several data portability implementations of the past.

2.1. Analysis of Rights

It is important to have the same understanding of terms for the analysis of rights. Thus, the terms stakeholder, right, obligation, delegation, rule statement, constraints and normative phrases are defined. For the definition of those terms this thesis focuses on Breaux [21]:

- A *stakeholder* is an entity afforded rights and/or obligations by Privacy Rule.
- A *right* is an action that a stakeholder is conditionally permitted to perform.
- An *obligation* is an action that a stakeholder is conditionally required to perform.
- A *delegation* is a right or obligation that a policy or stakeholder assigns to another stakeholder.
- A *rule statement* is the regulation text that includes the right or obligation and any constraints.
- A *constraint phrase* is part of a rule statement that describes a single precondition.
- A *normative phrase* contains words that indicate what "ought to be" as rights or obligations.

Semantic parameterization is a methodology for requirement analysis of regulations and policies [2] [1]. This procedure extracts discrete activities from rights and obligations by restating them in restricted natural language statements (RNLS). Those are then mapped to semantic models for further analysis. Breaux [21] applies this methodology to three different use cases - one of them is the Health Insurance Portability and Accountability Act (HIPAA). The methodology first conducts a policy goal mining from documents, before it applies the semantic parameterization [2]. The following paragraph describes the procedure:

The first step is goal mining from policies or documents [22]. Therefore, it is important to use action keywords within the goal as it describes what the actor has to do. In summary, the goal needs the elements *actor, action word*, *subject type* and *conditions*. An example can be the following:

> "Privacy statement:
> We share information for the purpose of marketing services.
>
> Goal: (Provider, SHARE information to market services). " [22]

In this example, the actor is the "provider", the action word is "SHARE", the subject type is "information" and the condition is "market services". Further examples of good action keywords are given by [1] which prevent ambiguity and encourage good understanding:

ACCESS	CONSOLIDATE	ENFORCE	MAXIMIZE	PROVIDE
AGGREGATE	CONTACT	ENSURE	MINIMIZE	RECOMMEND
ALLOW	CONTRACT	EXCHANGE	MONITOR	REQUEST
APPLY	CUSTOMIZE	HELP	NOTIFY	REQUIRE
AVOID	DENY	HONOR	OBLIGATE	RESERVE
BLOCK	DESTROY	IMPLY	OPT-IN	REVIEW
CHANGE	DISALLOW	INFORM	OPT-OUT	SHARE
CHOOSE	DISCIPLINE	INVESTIGATE	POST	SPECIFY
COLLECT	DISCLAIM	LIMIT	PREVENT	STORE
COMPLY	DISCLOSE	MAINTAIN	PROHIBIT	UPDATE
CONNECT	DISPLAY	MAKE	PROTECT	URGE

Following the goal definition, the next step is the semantic parameterization which starts with the derivation of restricted natural language statements (RNLS) [22]. RNLS focus on one activity which can reference to other RNLS. The activity describes how to fulfill obligations or right's details. An example for the derivation of RNLS is given below:

> " Goal: (Provider, SHARE information to market services).
>
> RNLS 1.1: The provider markets services.
>
> RNLS 1.2: The provider may share information to (RNLS1.1)." [22]

After determining the RNLS, the semantic models are structuring the statement meanings [22]. Therefore, first it is necessary to identify whether the RNLS is an obligation or a right and who owns this obligation or right. This is known as *root relation* which describes the main idea or concept. Further, the *associative relation* shows the conceptual relation to the root and includes the definition of the actor, action, object and target or purpose or instrument. The target, purpose or instrument can refer to another activity as well. Lastly, the *conceptual relation* maps the values derived from the RNLS to the conceptual relation. The following extract shows an example for the semantic model [22]:

```
activity[right:provider]{
    actor = provider
    action = share
    object = information
    target = ?whom
}
```

Due to the structured procedure and application to three policy regulations by Breaux (e.g. HIPAA [21]), this methodology is chosen for the analysis of requirements to comply with data portability GDPR Art. 20 [16]. The exact procedure is described in the next section.

2.2. Data Portability Requirement Analysis

This section describes the requirement analysis and its results of GDPR Art. 20. Before focusing on the detailed requirement analysis, the article and its structure is presented.

Overall, data portability according to GDPR Art. 20 contains two main roles: the data subject and the data controller. The data subject (DS) is a person who provides its data to the data controller. The data controller (DC) is a person or company who processes the personal data provided by the data subject.

The policy of data portability is structured into four paragraphs [16]:

1. *The data subject shall have the right to receive the personal data concerning him or her, which he or she has provided to a controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to another controller without hindrance from the controller to which the personal data have been provided, where:*

   (a) the processing is based on consent pursuant to point (a) of Article 6(1) or point (a) of Article 9(2) or on a contract pursuant to point

   (b) of Article 6(1); and (b) the processing is carried out by automated means.

2. *In exercising his or her right to data portability pursuant to paragraph 1, the data subject shall have the right to have the personal data transmitted directly from one controller to another, where technically feasible.*

3. *The exercise of the right referred to in paragraph 1 of this Article shall be without prejudice to Article 17. That right shall not apply to processing necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.*

4. *The right referred to in paragraph 1 shall not adversely affect the rights and freedoms of others.*

The paragraphs of article 20 refer to three other articles of the GDPR. They refer to article 6 paragraph 1 a and b where the topic of **"Lawfulness of processing"** is covered [16]:

1. *Processing shall be lawful only if and to the extent that at least one of the following applies:*

   (a) the data subject has given consent to the processing of his or her personal data for one or more specific purposes;

   (b) processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract;

Another reference by GDPR Art. 20 is article 9 paragraph 1 and 2 a, which discusses the topic of **"Processing of special categories of personal data"** [16]:

1. *Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation shall be prohibited.*
2. *Paragraph 1 shall not apply if one of the following applies:*

   (a) the data subject has given explicit consent to the processing of those personal data for one or more specified purposes, except where Union or Member State law provide that the prohibition referred to in paragraph 1 may not be lifted by the data subject;

The third and last reference of GDPR Art. 20 is the article 17 **"Right to erasure (right to be forgotten)"** [16]:

1. *The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:*

   (a) the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;

   (b) the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;

   (c) the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);

   (d) the personal data have been unlawfully processed;

   (e) the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;

   (f) the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).

2. *Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.*

3. *Paragraphs 1 and 2 shall not apply to the extent that processing is necessary:*

   (a) for exercising the right of freedom of expression and information;

   (b) for compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;

   (c) for reasons of public interest in the area of public health in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3);

   (d) for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or

   (e) for the establishment, exercise or defence of legal claims.

For the requirement analysis the data portability article itself, as well as its references are relevant. Thus, those will be the scope for the requirement analysis. The first step is the goal mining procedure. Further, the identified goals are rewritten into RNLS and finalized by the definition of the semantic models [22].

2.2.1. Goal mining of data portability GDPR Art. 20

The goal mining process for data portability GDPR Art. 20 concludes with the identification of goals. Those are displayed in table 2.1.

As some of the goals refer to other articles of the GDPR, the goal mining process has been conducted for the other articles as well. The results of the referenced articles 6 (1) a,b; 9 (1) (2) a and 17 are attached in appendix A. After deriving the goals from the policies, the next step is the identification of RNLS.

2.2.2. RNLS of data portability GDPR Art. 20

This subsection identifies RNLS for data portability according to GDPR Art. 20. The resulting RNLS for GDPR Art. 20 are summarized in table 2.2.

Actor	Action Word	Subject Type	Conditions / Reference
DS	RECEIVE	personal data	concerning him or her
DS	RECEIVE	personal data	has provided to controller
DC	PROVIDE	personal data	concerning DS
DC	PROVIDE	personal data	provided by DS
DC	PROVIDE	personal data	structured, commonly used and machine-readable format
DS	TRANSMIT	personal data	to another controller
DS	TRANSMIT	personal data	without hindrance from the controller
DS	GIVES	consent pursuant	to point (a) of Article 6 (1)
DS	GIVES	consent pursuant	to point (a) of Article 9 (2)
DS	GIVES	contract pursuant	to point (b) of Article 6 (1)
DC	PROCESS	personal data	automated
DC	TRANSMIT	personal data	directly to another controller where technically feasible
DS	MAINTAIN	right	Article 17
DC	DENY	data deletion	for necessary data for performance of a task carried out in public interest
DC	DENY	data deletion	in exercise of official authority vested in controller
DS	ENSURES	rights and freedoms	of others

Table 2.1.: Goal mining of data portability GDPR Art. 20.

RNLS#	Ref#	Statement
1		DS provide personal data concerning him or her.
2	1	DC provides stored personal data to RNLS#1
3	1	DC provides stored personal data in structured, commonly used and machine-readable format to RNLS#1.
4		DS transmit personal data to another controller.
5	4	DC does not hindrance transmission RNLS#4.
6	15	DS gives consent pursuant to processing personal data (RNLS#15).
7	19	DS gives consent pursuant to processing special categories of personal data (RNLS#19).
8	16, 17	DS gives contract pursuant to processing personal data (RNLS#16, RNLS#17).
9		DC process personal data automated.
10	1	DC transmit personal data from RNLS#1 directly to another controller where technical feasible.
11	21	DS maintain right to Article 17 (RNLS#21 ongoing).
12	21	DC deny RNLS#21 if necessary for performance of a task carried out in public interest or in exercise of official authority vested in controller.
13		DS ensures rights and freedom of others.

Table 2.2.: RNLS of data portability GDPR Art. 20.

The RNLS definition for the references within the GDPR, articles 6 (1) a, b; 9 (1) (2) a and 17, are listed in appendix B. A graph for the visualization of RNLS references is displayed in figure 2.1. This figure should show the relations and influence of Art. 20, Art. 6, Art. 9 and Art. 17.

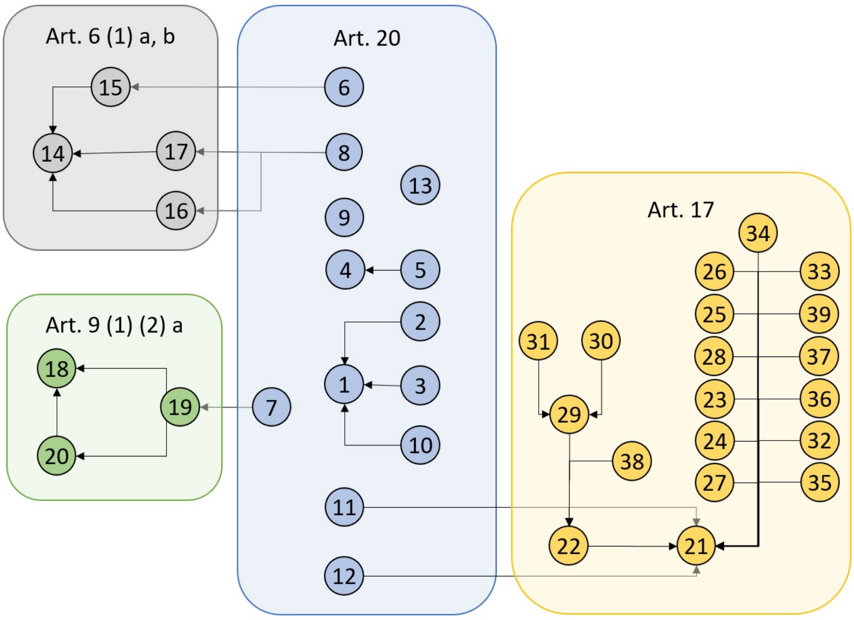

Figure 2.1.: RNLS relations as a graph with nodes.

2.2.3. Semantic models of data portability GDPR Art. 20

The final step of semantic parameterization [1] [22] [2] [21] is the definition of a semantic model for the RNLS. The semantic models for data portability are described in table 2.3. Appendix C displays the semantic models of referenced articles 6 (1) a, b; 9 (1) (2) a and 17.

2.2.4. Necessary properties according to data portability GDPR Art. 20

The identification of the semantic models leads to the question, which information is necessary to answer the semantic models. This means, the properties about the requester need to be found within the company. The overall process from Breaux [2] [1] and the additional analysis are displayed in figure 2.2.

Figure 2.2.: Abstract process of rights analysis [1] [2] and additional assessment.

Semantic Models						
#	**activity**	**actor**	**action**	**object**	**target**	**purpose / instrument**
1	right: DS	DS	provide	personal data	concerning him or her	
2	right: DS	DS	provide	personal data	to DC	
3	obligation: DC	DC	provide	stored personal data	DS	
4	obligation: DC	DC	provide	stored personal data		structured, commonly used, machine-readable format
5	right: DS	DS	transmit	personal data	another controller	
6	obligation: DC	DC	doesn't hinder-ance	transmission of personal data	by DS	
7	right: DS	DS	give	consent pursuant	DC	for processing for one or more purposes
8	right: DS	DS	give	explicit consent pursuant	DC	processing special categories of personal data
9	right: DS	DS	give	contract pursuant	DC	enable DS to enter contract
10	obligation: DC	DC	process	personal data		automated
11	obligation: DC	DC	transmit directly	personal data	another controller	where technically feasible
12	right: DS	DS	maintain	rights to erasure (Art. 17)	DC	
13	right: DC	DC	deny erasure	stored personal data	DS	performance of carried out task in public interest / in exercise of official authority vested in controller
14	right: DS	DS	ensure	rights and freedom	others	

Table 2.3.: Semantic models of data portability GDPR Art. 20.

As a result, there are seven properties necessary to fulfill the requirements of data portability according to GDPR Art. 20. Those are listed in table 2.4. After identifying the required information for the semantic models, those requirements are summarized as properties. As already mentioned in the last sections, GDPR Art. 20 references other articles within the regulation. The properties displayed in the table also contain a reference to the respective article, which states the root article and necessity of the property. This shows, that most information required for the fulfillment of data portability is a summary of information according to the other referenced articles. The information model development should consider this list of properties.

Information Property	**Article**
personal data concerning the data subject	data portability (Art. 20)
consent pursuant	Lawfulness of processing (Art. 6 (1 a, b))
explicit consent pursuant	Processing of special categories of personal data (Art. 9 (2 a))
specific categories of personal data	Processing of special categories of personal data (Art. 9 (2 a))
contract pursuant	data portability (Art. 20)
erasure date	Right to Erasure (Art. 17)
public interest	Right to Erasure (Art. 17)

Table 2.4.: Information properties for requirement fulfillment.

2.3. Data Portability Challenges

Data portability, as described in GDPR Art. 20, leads to implementation and interpretation challenges for companies. This chapter discusses the main issues that appear when trying to fulfill the regulation's requirements. First, it addresses limitations of data portability entitlement. Second, the similarities and differences of data portability according to GDPR Art. 20 and other legal foundations are reviewed.

2.3.1. Data portability limitations

The conference paper of Bistolfi et al. [23] discusses one proposal of the GDPR [24] from 2015. The paper reveals that there already were several issues with the interpretation and implementation of GDPR Art. 20 before the agreement on the final regulation [16] was published in 2016. Also, [25] criticizes several aspects of a draft version of the GDPR from 2012 [26].

One early identified problem by [23] is the missing availability of standards which would enable the interoperability of systems. This leads to an additional obligation for data controllers to develop and foster a standard for data portability. Subsequently, market leaders may be able to influence the development of standards and make it difficult for competitors to adapt. Even though the commission proposed to use governmental acts to encourage the determination of technical standards, the proposal was not considered in the final regulation [25]. Another issue is the transfer of personal data from one controller to another where it is "technically feasible". This may lead to a security problem if the other controller does not ensure at least the same security level [23]. Thus, the legal feasibility of the direct transfer should also be considered. Similar to this legal issue, the data controller should keep in mind that the right for data portability only applies if the data subject is authenticated and authorized [23].

In summary, the critical discussion of the proposal for data portability concludes that the regulation is *"more a declaration of principle than a real and effective tool for individual self-determination in the digital environment."* [23] Most of the issues identified in the GDPR proposal remained after its publication.

According to [27], one of the key issues is the limited access to data which is generated by the data controller. Data portability complying with GDPR Art. 20 does not cover the transfer of data which is generated by data controllers for internal, statistical or analytic purposes [28]. GDPR Art. 20 states that for the fulfillment of requirements only data provided by the data subject themselves is within scope. This means that even if data generated by the data controller or other persons is available at the IT service, it may not be part of the transferred data as it is not provided by the data subject. An example given here is the ratings of a retailer on a website [27]. These are not included in the transferable dataset as they are generated by feedback from other users and calculation by the data controller themselves.

The next issue with data portability according to GDPR Art. 20 concerns privacy rights of third parties [27]. The data controller can deny a request for data transfer if the rights or freedoms of others are affected. An example given by Vanberg et al. [27] is the transfer of pictures to another platform which can be denied if other people appear in that picture. As this hampers the transferability of data, data subjects may be discouraged in utilizing their right. Furthermore, the other person who appears in the picture can prohibit the transfer of the data [29]. Nevertheless, there is no impairment of rights of third parties when transferred data is used for the same purpose such as by the previous data controller [30].

As already described by [23], the term "technically feasible" causes confusion and obscurity. Also, [27] discusses this problem by associating the issue with the problem of missing standards, which is the main focus of [23]. According to [27], the GDPR is missing detailed information about what exactly is the necessary electronic format for fulfillment of "technically feasible". As every data controller may have their own understanding of this wording, there is no general comprehension. Thus, data controllers may claim that the direct data portability to another controller may not be technically feasible. Furthermore, the issue with the missing information details about the electronic format leads to the necessity to develop standards within industries [23] [27]. The issue with the unclear meaning of technically feasible is as well discussed by [25]. They discuss the missing definition by stating that the data controller is not responsible to set up compatible data processing systems, but to enable data transfer with existing possibilities. Moreover, the data controller should enable an adaption to an electronic format in relative extent. But the meaning of "relative extent" is again not defined.

When it comes to the implementation of data portability, just like in any other change project, the costs and efforts are in focus. To enable the data transfer to other data controllers, many companies have to spend a high amount on adaption costs [27]. The GDPR is not limited to any costs of implementation, which depend on the company's size. To comply with GDPR Art. 20, especially small and medium-sized companies are affected negatively with respect to the costs for implementation [27]. The issue with costs and efforts was pointed out in December 2015 by Baroness Neville Rolfe, the UK's parliamentary Under-Secretary of State for the Department for Business, Innovation and Skills, who stated that enabling data portability should not imply too high costs as it may constitute a barrier for market entry [31]. Thus, this may lead to negative effects for innovation and a level playing field contrary to the objectives of the data portability right of GDPR Art. 20 [31]. In contrast to the challenges regarding costs and efforts stated by [27] [31], the Impact Assessment by the European Commission [26] concluded in 2012 that the adaptation of systems to enable data portability implies minimal costs, as there are already existing interfaces such as for third party developers. According to [25], the estimation by the European Commission regarding the costs and efforts for the implementation of data portability is not yet confirmed, as GDPR Art. 20 involves various data types and stakeholders. Thus, the extent of implementation is still unknown.

Another topic to consider regarding data portability is the transfer of data that contains valuable proprietary data or intellectual property [27]. Many evaluations [29] [23] [27] point out that GDPR Art. 20 can be seen as an extension to the *"Right of access by the data subject"* GDPR Art. 15. Hence the following question has to be cleared: What are the differences to consider between Art. 15 and GDPR Art. 20? According to the Recital 63 of the GDPR, the right of access by the data subject contains a limitation. This limitation describes the reduction of the personal data export if and only if rights and freedoms of others including trade secrets or intellectual right properties are affected [27]. Neither the GDPR, nor the Recital 63 contains a specific reference to Art. 15 or any information about a restriction for proprietary data or intellectual property. Thus a clarification of interpretation is necessary [27].

If a data subject wants to enforce their right to data portability, they will face obstacles in the sense of missing information on how to enforce their rights such as data portability [27]. The Article 29 Working Party already mentioned this issue within their recommendations for implementation of the GDPR in 2016 [32]. Consequently, a guideline and explanation shall be provided to the data subject by the data controller which explains how the data subject may enforce the right of data portability correctly. This is not explicitly an obligation according to the GDPR Art. 20. Nevertheless, it is recommended [32] [27] as otherwise it might discourage data subjects in enforcing their rights.

Finally, another challenge in complying with the GDPR Art. 20 are new possibilities of security and privacy breaches [27]. The main objective in data portability is to enable interoperability between companies. According to [33], interoperability is one of the main factors with respect to opportunities for security breaches. This means that by enabling data transfer to another controller, the data controller needs to consider additional security risks while fulfilling the obligation [27]. This implies further duties like securing data transfer itself as well as providing proper authentication and authorization of the data subject, who requests the data transfer [27].

2.3.2. Comparison of data portability (GDPR Art. 20) and other legal basis

This section discusses different legal bases for data portability according to the GDPR Art. 20 and other regulations.

As already mentioned above, data portability GDPR Art. 20 may be seen as an extension to Art. 15 GDPR - the *"Right of access by the data subject"* [29] [23] [27] [25]. Both articles contain the obligation to provide a digital extract of personal data which are stored about the data subject. The difference between those two articles is the format. While Art. 15 demands a human understandable, Art. 20 claims a machine-readable format [29]. This distinction has to be considered during implementation.

As the main objective of data portability GDPR Art. 20 is to foster a level playing field, the similarities and differences to the European Union Competition Rules [34] should be considered. The major difference between those two regulations is that the EU Competition Rules [34] apply to only dominant market players while the GDPR applies to all companies on the market, irrespective of size or market share [27]. In summary, the GDPR intents to lower the barrier for market access, avoid high switching costs and reduce network effects that may hamper competition [27].

As Vanberg et al. [27] summarizes, the implementation of the GDPR should not be in conflict with competition laws, consumer protection laws or others. Thus, a cooperation is necessary between several stakeholders who may have a conflict with GDPR implementations e.g. competition authorities, European Data Protection Supervisor, national data protection agencies etc.

2.4. Data Portability Implementations

With the objective of developing an information model that fulfills the requirements of data portability GDPR Art. 20, the question arises, whether there are existing similar implementations for data portability. For this purpose, the following section discusses implementations covering data portability in a similar way.

2.4.1. Data portability according to HIPAA

When reviewing the topic of data portability within literature, several laws for specific industries introduced similar functionality requirements in the past. One example is the American HIPAA of 1996 [35]. These privacy, security and breach notification rules entitle individuals to their health information [36]. For example, the individual has the right to receive health plans and a health information copy from the health care provider. Furthermore, it covers the topic of electronic exchange of health information by using a HIPAA electronic transactions and code set standard. The HIPAA privacy rule contains an additional description of data portability for health records [36]:

- individual has to request right enforcement
- individual receive copies of information in medical or health records provided and managed by HIPAA covered entity (health care provider or health plan)

- transmission of health information must take within 30 days either directly to a person or other covered entity

Besides, enabling data portability to individuals and other health care providers, HIPAA may be used for sharing health information with operating services like treatment, payment and health care operations [36]. Health IT fosters in further improvement of the exchange of health information. In 2015, they developed a standard for the view, download and transmission of health records (Certified Electronic Health Record Technology) by providing Application Programming Interfaces (API) [36]. At the moment, HIPAA adopts standards for electronic transactions for the purpose of better efficiency and effectiveness of the U.S. health care system [37]. According to [37], the described standards apply to all HIPAA covered entities, which are health plans, health care clearinghouses and health care providers. Currently, a 21st Century Cure Act is under development, which should strengthen the interoperability of health information [36].

2.4.2. Data portability according to the Payment Accounts Act

Back in 2014, the European Parliament announced the Payment Accounts Act which implies new rights for bank customers [38]. In summary, the new law provides a basic payment account, fee transparency and help with switching bank accounts [39]. In Germany, the European Payment Account Directive has been implemented by the Federal Financial Supervisory Authority and became effective in 2016.

Just like data portability (GDPR Art. 20), the Payment Accounts Act's intention is to foster a level playing field by enabling an easier bank account switching process [39]. In detail, the Payment Accounts Act requires enforcement by the customer as well. The responsibility for the bank are to enable an uncomplicated transfer of banking details. This obligation is binding for the receiving and transferring bank, who are in joint security.

The obligations for the receiving and transferring bank are the following [40]:

The receiving bank has to request a list about existing information about standing orders and SEPA Direct Debit Mandates which shall be transferred. The transferring bank should provide the list within five working days. Furthermore, the transferring bank has to transfer the current account balance and close the account at the desired time. After receiving the information about standing order and SEPA Direct Debit Mandates, the receiving bank has to set up the bank account properly within five working days.

So far, the technical implementation of the account switching procedure is not yet described in literature.

2.4.3. Data portability according to the German Telecommunication Act §46

Another interesting directive published in 2012 is the Telecommunication Act §46 [41]. It covers the topic of the portability of phone numbers in Germany. With this law, customers get the right to keep their phone number in case they want to switch service provider [42].

To enforce the right of phone number transfer, the customer has to request the transfer at the new service provider [42]. To pave the way for a proper transfer of the phone number, it is important that customer data stored at the old and new service providers are structured identically [42]. Thus, customers should update their customers' data at the old service provider before the transfer takes place [42].

The data structure of a customer's data is the following for private customers:

- Phone Number
- Name
- Date of Birth

Reflecting the Telecommunication Act §46, the question arises whether data portability according to GDPR Art. 20 distinguishes from this directive. According to [23], data portability in GDPR has a much broader scope. Furthermore, the GDPR limits data portability to the situations where it is "technically feasible". In the Telecommunication Act §46, the portability of phone numbers is always managed electronically. Thus, the GDPR does not specify a standard which as well differs from the Telecommunication Act §46 [23].

3. Information Model

For the development of an information model it is important to know what the basis for the information model should be. For this purpose, the term *"information model"* is defined and analyzed by different perspectives within the first subsection. After understanding the information model term, different abstraction layers should be discussed in the following subsections. Within the discussion of information model approaches, the method for modelling based on documents is analyzed. Moreover, another abstract view for the development of an information model on the business object layer is discussed. Each approach is analyzed by emphasizing the advantages and pointing out the disadvantages of the concept.

3.1. Information Model Basics

This part discusses the basics for information models. It contains the definitions of a model, information model and related terms. Moreover, the modelling of information models and possible evaluation methods are discussed.

3.1.1. Model and information model definitions

Prior to the definition of the term information model, the term *model* already has multiple definitions in theory. The International Electrotechnical Commission defines a model as a:

> "mathematical or physical representation of a system or a process, based with sufficient precision upon known laws, identification or specified suppositions." [43]

For the purpose of the model within this thesis, the definition of [43] is more focused on the mathematical and physical description of a model, but not on an information system view.

Another definition of a model is given by Stachowiak:

> "A model is a restrict depiction of the reality." [44]

Moreover, [44] describes three characteristics of a model. The first is the depiction of natural or artificial originals. The second is the reduction of the original and the depiction of only relevant attributes. The third characteristic is pragmatism, which describes the replacement function of a model.

A further definition is given by Kastens et al. which basically defines a model as

> "an extract of the original." [45]

Also, [45] describes the process of modelling to select only relevant aspects of the original as the complete modelling of the original is not advisable. Furthermore, the purpose of modelling is to visualize, understand, draft, simulate, analyze or communicate about a system [45]. According to [45], a model contains a structure of the original, characteristics of the original, relations between parts of the original and behaviour of the original. Reflecting the definition of [45] and [44], there are several similarities, such as the focusing on only relevant aspects of the original. Therefore, this thesis will use the definition by [45] as a basis for the term "model".

After the definition of the term "model", more specific definitions of "information model" are discussed: [14, p.148] describes an information model as a model which represents the structure and semantics of the information from a conceptual view. The semantics describe the conditions necessary to assure consistent information. Further, the semantics show which information is used and its correlation to other information. Moreover, [14] defines the information model as a meta model, which stores metadata. For instance, a data model is stored but not the actual data. Another example is the storing of information about programs but not the actual programs.

With respect to [14, p.175], the information model elements are the systematic and structured description of information systems. With the help of information systems, existing information and communication structures are displayed. Information models describe an electronical information processing. In more detail, they are a system to process, create and transfer information.

The definition of information model used within this thesis is given by [3, p.63]: An information model is a specific model focused on systems and mainly used within the field of information systems. An information model is the result of a construction by a modeler [3]. This construction is the result of the information modelling of system elements which are seen as essential by the modeler. Another characteristic of an information model is that it is a description model as it usually does not contain any decisional character [3]. Moreover, the information model can be categorized into two applications: a single use case, e.g. a specific use case within a company, or a group of use cases, e.g. an information model for a specific problem which is not restricted to one specific company [3, p.66]. The latter information model can also be called reference model and is a general description of a problem by a domain [3]. This reference model has to be adapted by each company for their environment, which is the main difference from a company specific information model.

Besides the differentiation of information models and its usage for one specific or a group of companies, the abstraction level is also an important aspect of an information level [3]. Schütte describes three different levels: the *instance*, *type* and *abstract* level. While models on instance level are duplicates of the original, models on type level focus on classes of elements [3]. The higher abstract level is obtained by orchestrating classes of the type level to clusters [3]. The definition of clusters is based on contextualized criteria, which depends on the use case of the information model.

A further distinction of information models by model language is possible [3]: object language, meta language, meta-meta language. The basis for differentiation is given by the context in which the language is used. Thus, meta language describes an object language. This distinction is used to differentiate between object and metamodel. While an object model represents an object, a metamodel describes another model e.g. an object model.

Usually, the syntax of a language is used to describe a metamodel.

In detail, the elements of an object model, meta model and meta-meta model can be described as the following [46]:

- Meta-meta model: Modelling meta objects, relationships between meta objects (e.g. Classes, Associations, Data Types) (e.g. Class Person can be either Customer or Employee, Associations: A Person can be related to a Product)
- Metamodel: description of models (e.g. UML-Metamodel, Entity-Relationship-Metamodel) (e.g. Entity Customer is of class Person and has relation to Entity Invoice)
- Object model: description of data or objects (e.g. UML-Statechart-Diagram, Entity-Relationship-Diagram) (e.g. Customer buying product)
- Data or Objects: Instances of objects, data fields (e.g. Customer1: Smith)

Besides considering the abstraction level of an information model, there are three additional characteristics to take into account in the process of information modelling [46]:

- The Accuracy of the information model considering the defined syntax.
- The Integrity of the information model considering all relevant information object types and their relations.
- The Consistency of the information model considering the complete specification of the used metamodel [3, p.126].

In order to enable automation of the processing of an information model, metamodel information is necessary as semantics of the model are of higher relevance [46]. An essential distinction has to be drawn between the terms “reference model” and “metamodel” [3, p.72]. A reference model describes the semantics of a model by abstracting its syntax. A metamodel describes the syntax of a model and is an abstraction of the semantics. The combination of a reference model and metamodel is a reference model of a metamodel which describes the similarities of metamodels in a reference model. This relationships between the different models can be seen in figure 3.1.

Finally, for the process of information modelling, the syntax of the model language as well as the semantic reference is necessary [3, p.75]. With this background, the next question is how information modelling should be done.

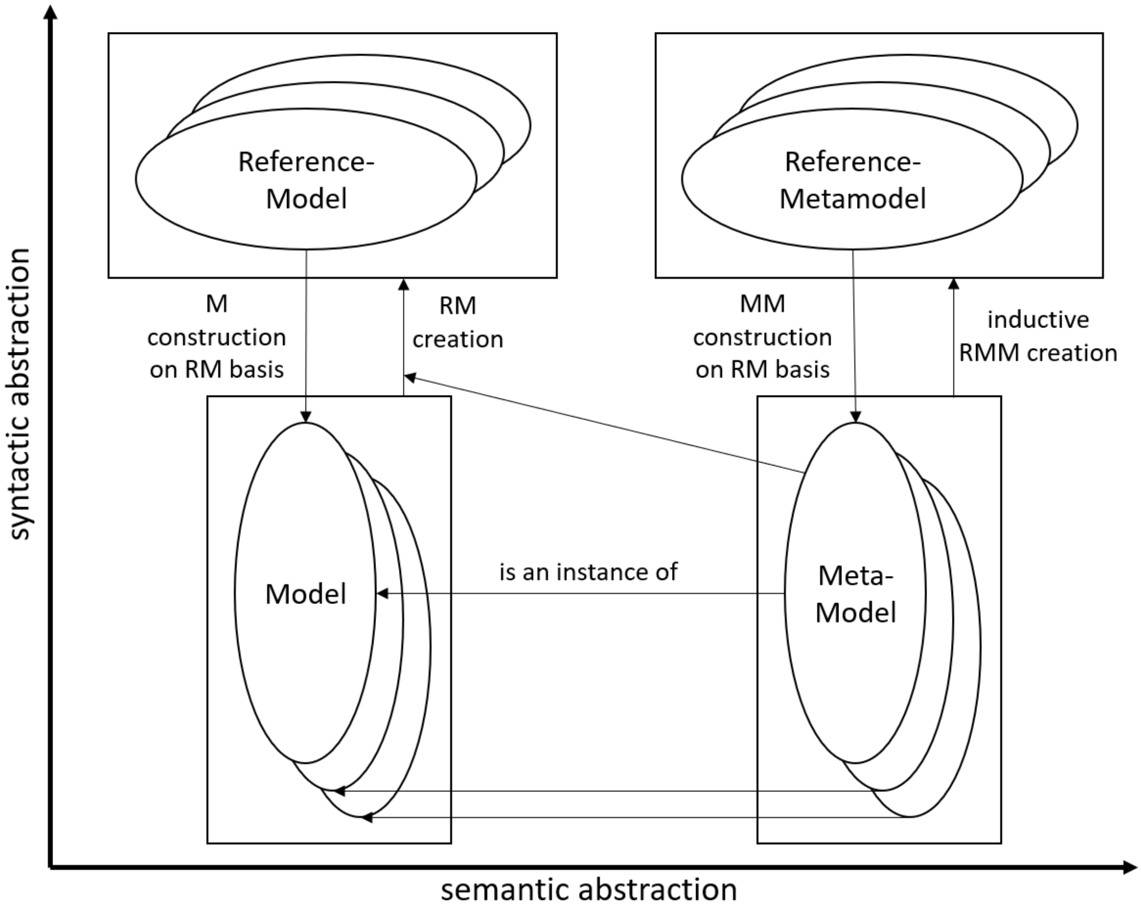

Figure 3.1.: Distinction of reference model and metamodel [3, p.73].

3.1.2. Information modelling

This section analyzes the process of information modelling in detail. The information modelling process of [3, p.184] describes how reference models are designed. The process contains five phases in total and is visualized in figure 3.2.

The first phase starts with the problem definition. Therefore, input by several stakeholders concerned by the issue is necessary. They can define the problem situation and are familiar with the field in question. The result of this phase is a hypothesis of problems which seems to be most relevant for the problem domain. After the definition of the problem, a top-down modelling can begin to assure the highest probability of a best-fitting information model for the problem domain.

The second phase is about the construction of a model frame. The classification of elements within and characteristics of the model will answer the question *what* a model needs to define. Therefore, a master-model should be used, which already contains standardized terms and model elements. The result is the differentiation of configuration relevant or irrelevant knowledge and the connection between a structural and behavioural perspective.

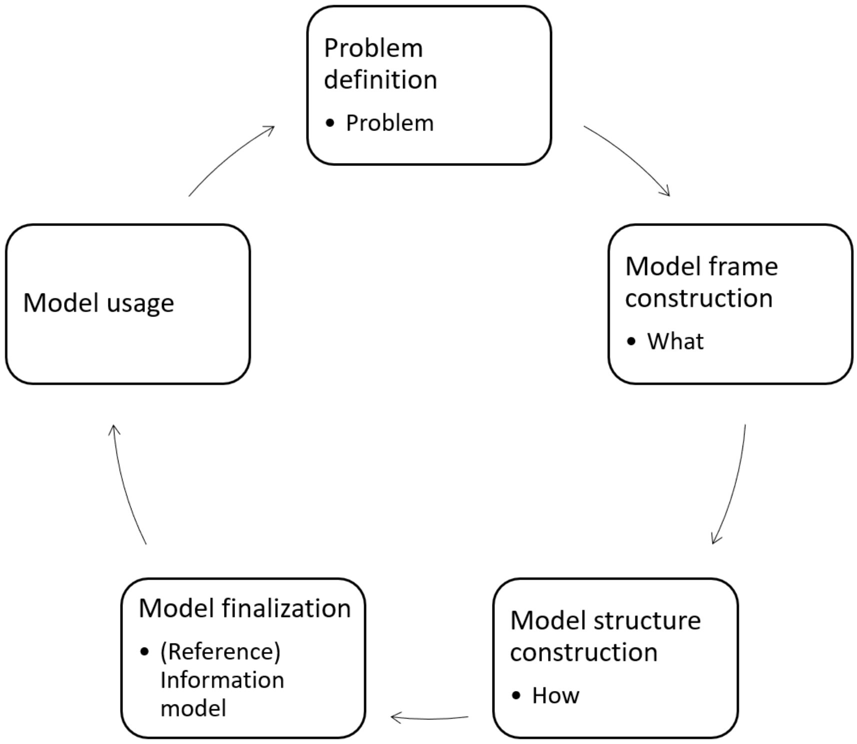

Figure 3.2.: Process for information modelling [3, p.185].

The third phase is the construction of a model structure. After the model elements identification in the previous phase, the next step is to focus on more details such as defining *how* the model can be described. For this purpose, a formal language and a model structure is necessary. Modelling languages contain a structure which is described by the syntax. The language syntax enables a quality check by stakeholders against common rules given by the used modelling language.

The fourth phase is the finalization of the model. Within this phase, relations between elements and other models are identified. This improves the consistency of the model. Furthermore, additional quantitative statements about the model are required to enable the model's quality checking and benchmarking.

The fifth and last phase is the implementation and model usage. The implementation and usage of a model depends on individual and company specific factors, which are hard to standardize and describe in general. General reference models can only be valuable if it is applied to a specific use case. Thus, the construction and the usage of a reference model can be used as basis for the development of a company specific model. The difference between a reference and a modified model can identify potential extensions to the reference model.

Those five phases are good approaches in order to focus on essential aspects of an information model and to assure the compliance with modelling principles or other quality criteria which are elements of information model evaluations (3.1.3).

Referring to the problem definition, the selection of input data used for the information modelling process is important. It can be categorized in different knowledge sources. According to [14], those sources can be provided by:

- Experts
- Documents
- Forms
- Software
- Formal Model

Each knowledge source has different advantages and disadvantages. For information modelling the criteria of consistency, problem fit, effort, information density, integration potential, generality and completeness are essential [14]. Table 3.1 displays the knowledge source and their usage potential for information modelling which is evaluated as either good, medium or bad.

Criteria / **Knowledge Source**	**Consistency**	**Problem fit**	**Effort**	**Information density**	**Integration potential**	**Generality**	**Completeness**
Experts	bad	good	bad	medium	good	bad	good
Documents	good	bad	medium	bad	good	good	good
Forms	good	good	good	good	medium	bad	bad
Software	bad	medium	bad	bad	bad	medium	bad
Formal Model	good		good	good	good	good	medium

Table 3.1.: Knowledge sources for information modelling [14].

3.1.3. Information model evaluation

Besides the development, the quality evaluation of an information model is interesting as well. For this purpose, two approaches for assuring the quality of an information model [3] [47] are discussed within this subsection.

The evaluation by Schütte [3] describes ”Grundsätze ordnungsmäßiger Modellierung”, meaning ”Modelling Principles”, and contains six criteria which evaluate the quality of a model's construction.

The first modelling principle is the "principle of design adequacy" which covers the topic of rationality of the model. In other words, it evaluates the usage of the model with respect to the problem domain.

The second modelling principle is the "principle of language adequancy" which evaluates the usage of language syntax and of the design language for the specific problem.

The third modelling principle is the "principle of efficiency", which handles the topic of economical restrictions e.g. costs and savings by using the information model.

The fourth modelling principle is the "principle of systematical structure". It evaluates the information model whether it follows a systematic structure consistently throughout diverse view points and the problem domain.

The fifth modelling principle is the "principle of transparency". This principle focuses on the clarity of the information model and examines the same understanding for every stakeholder.

The last and sixth principle is the "principle of comparability" which contains the evaluation of the comparability and integration of one information model to other information models. Moreover, this principle assesses the transferability of the information model to others.

In summary, table 3.2 shows the modelling principles, their objectives and whether the evaluation is from a objective or subjective perspective.

Modelling principles	**Objectives**	**Evaluation type**
Principle of design adequacy	Consensus on problem definition, model representation	subjective
Principle of language adequacy	linguistic correctness (consistency, completeness) and language suitability (e.g. semantic power, comprehensibility and usage of the language)	objective and subjective
Principle of efficiency	Consensus finding, Language understanding and usage, translatability, across views	subjective
Principle of systematical structure	information system architectures, inter-model relations	objective
Principle of transparency	hierarchical clarity, layout comprehensibility, filerting	subjective
Principle of comparability	semantical comparability	subjective

Table 3.2.: Modelling principles in Summary [3, p.136].

Besides the modelling principles of Schütte [3], there is another interesting approach by Lindland et al. [47]. This method evaluates the quality of information models as well and is based on sign theory, developed for conceptual models like data models, process models, interaction models etc. According to [47], the quality of a conceptual model can be described by four components:

- the model M with represented statements.
- the language L contains possible statements with respect to the used modelling language.
- the domain D handles the topic of how the specific problem is solved with the respective statements of the model.
- the interpretation I is measuring how good the stakeholders can interpret, use and understand the model.

Along with the components of the quality framework, the framework defines quality categories according to the three levels of semiotic theory (the theory of signs) [48]: Syntactic, semantic and pragmatic quality.

The syntactic quality of a model describes the realization of the modelling language used within the model. Thus, the model should only use statements which are *syntactically correct*.

Next, the semantic quality describes the *validity* and *completeness* of the model with respect to the problem domain. This means, there should be only statements in the model covering relevant or correct aspects of the problem description.

The third and last level is the pragmatic quality. With this quality category, the interpretation of the model by the stakeholders and the objective of *comprehensibility* is described.

In summary, the quality framework by Lindland et al. [47] and the semiotic theory levels by Morris et al. [48] are visualized in figure 3.3.

An evaluation of the proposed quality model of Lindland et al. [47] was conducted by Moody et al. who determines the framework as

> "valid, with the quality categories found to be necessary, sufficient and independent." [4]

Furthermore, the conducted experiments with 192 participants received positive feedback, especially for the high usability and its applicability for the evaluation of information models [4].

Comparing the information model evaluation methods of [3] and [47], the advantage of [47] outweigh. It is a more comprehensive approach and the positive outcome of the framework evaluation by [4] convinces. Unfortunately, the modelling principles by [3] have no scientific evaluation like the quality framework by [47]. Thus, Lindland's criteria should be used for the evaluation of the information model.

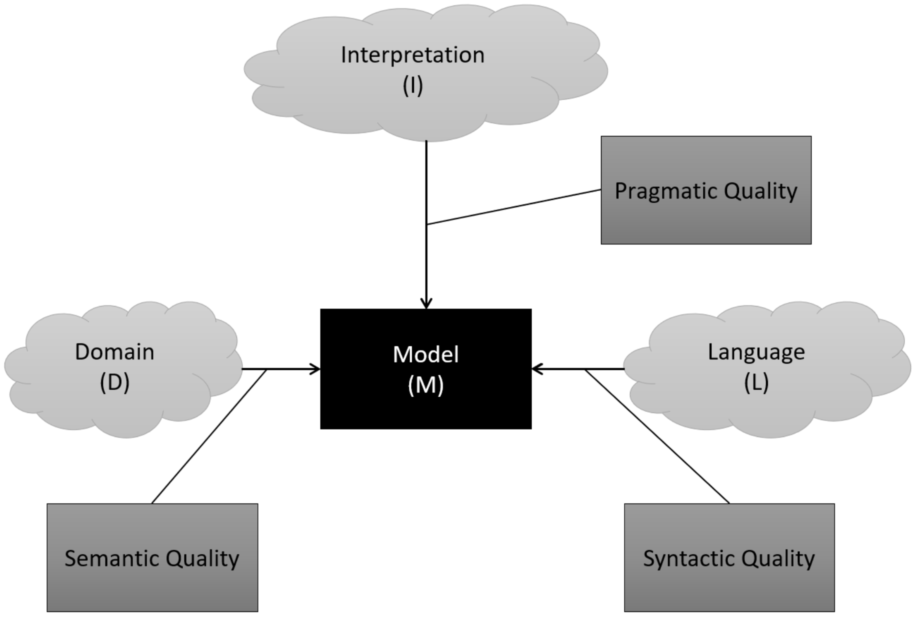

Figure 3.3.: Conceptual model quality framework [4].

3.2. Document-based Approach

One idea for the information model development is to focus on documents as a basis. Documents are for example files like invoices, customer databases or similar storage where personal data may be stored. To visualize the idea, figure 3.4 displays an abstract architecture.

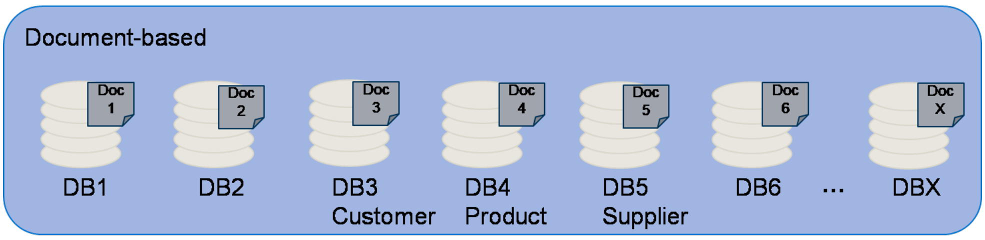

Figure 3.4.: Abstract example for database and documents as basis for information model.

For the document-based approach, a metadata model is one way of defining an information model. Before having a more detailed look at metadata models, it is important to know what metadata is. According to Gartner, metadata is defined as an

> "information that describes various facets of an information asset to improve its usability throughout its life cycle. It is metadata that turns information into an asset. Generally speaking, the more valuable the information asset, the more critical it is to manage the metadata about it, because it is the metadata definition that provides understanding that unlocks the value of data." [49]

Another definition by the National Information Standards Organization (NISO) is

> "metadata is data about data." [5]

For this thesis the definition by NISO is relevant. Thus, the NISO definition is described in more detail:

NISO distinguishes four diverse types of metadata: descriptive metadata, administrative metadata, structural metadata and markup languages.

The first type "descriptive metadata" is used to find or understand a resource with the use case of the discovery, the display and interoperability of data.

Second, the administrative metadata has three more subcategories "technical metadata", "preservation metadata" and "rights metadata". Technical metadata is used to decode and render files and is as well for the use cases of interoperability, digital object management and preservation of data. For managing long-term files, the preservation metadata is also used for the interoperability, digital object management and preservation. The rights metadata are intellectual property rights which are related to the content. Use cases for the rights metadata are the interoperability and the digital object management of data.

Third, to define the relationships of resources to each other, "structural metadata" may be used. This type of metadata is primary used for the navigation through data.

The last category of metadata is "markup languages" which uses the integration of metadata and flagging for other structural or semantic features within the content. Markup languages are used mainly for the navigation and interoperability of data.

There are different standards which are used in context of metadata and metadata models. The Resource Description Framework (RDF) contains World Wide Web Consortium (W3C) specifications which are designed for metadata of the semantic web [5]. Another metadata language is Dublin Core which contains 15 elements as metadata model for digital information [15]. Those two standards are presented in the next subsections.

3.2.1. RDF and the Semantic Web

The RDF models information and its relations as a graph [5]. This graph represents a network which can be accessed at any point. Elements of the RDF graph are described as a triple which consist of the *subject*, *predicate* and *object* [5]. In other words: the subject is displayed by a class; the predicate is described by properties and the object is a member of a specified class. Furthermore, the RDF graph can define properties within a domain and a range of properties [5]. In this case, a domain signifies that a subject is part of a specified class. A range describes the belonging of an object to a specified class. The triples are represented by Uniform Resource Identifiers (URIs) or International Resource Identifier (IRIs) [5]. As the definition of classes and predicates follows a structure, it enables referral of triples to each other. Further, the structure may be used to connect classes and properties to other classes or properties [5]. Figure 3.5 shows an example of an RDF graph.

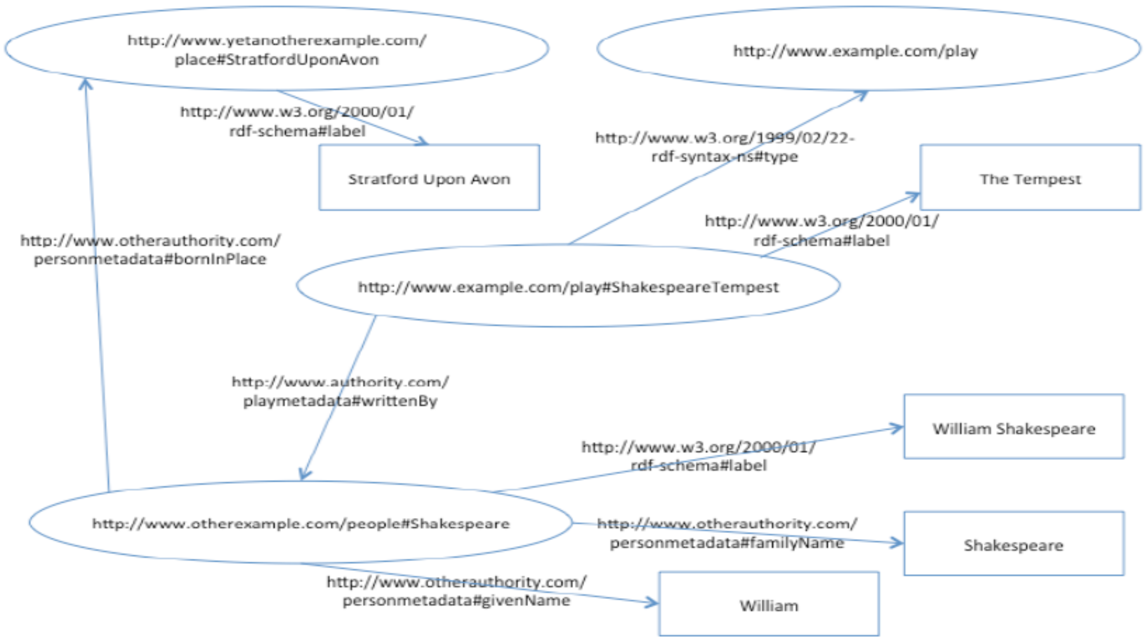

Figure 3.5.: Example for an RDF graph: A simple graph for Shakespeare's *The Tempest* [5].

RDF data has the main advantage of being highly compatible and is commonly used with HTML, which contains embedded structured metadata [5]. Moreover, there are several RDF syntaxes like RDF/XML or RDF within HTML which translates the RDF graph into their specific syntax [5]. An example for RDF/XML serialization is the following code [5]:

```
<?xml version = "1.0" encoding ="URUF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
    xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
    xmlns:ex="http://ww.example.com/"
    xmlns:play="http://www.authority.com/playmetadata#"
    xmlns:person="http://www.otherauthority.com/personmetadata#">
    <ex:play rdf:about="http://www.example.com/play#ShakespeareTempest">
        <rdfs:label>The Tempest</rdfs:label>
        <play:writtenBy>
            <rdf:Description
rdf:about="http://www.otherexample.com/people#Shakespeare">
                <rdfs:label>William Shakespeare</rdfs:label>
                <person:givenName>William</person:givenName>
                <person:familyName>Shakespeare</person:familyName>
                <person:bornInPlace>
                    <rdf:Description
rdf:about="http://www.yetanotherexample.com/place#StratfordUponAvon">
                        <rdfs:label>Stratford Upon Avon</rdfs:label>
                    </rdf:Description>
                </person:bornInPlace>
            </rdf:Description>
        </play:writtenBy>
    </ex:play>
</rdf:RDF>
```

Recapitulating, the RDF enables referencing triples and elements of the triple to each other which is an essential value for linked data. Linked data connects data and vocabularies from multiple sources and is an essential aspect of the vision for the Semantic Web: *"Building the Web of Data"* [5]. The Semantic Web initiative founder Tim Berners-Lee has the vision of a web which has the characteristics of automation, self-adaption, integration and flexibility [6]. Furthermore, the purpose of the Semantic Web is to be more focused on interaction with the user and to offer a better and more convenient experience in the web [6]. The term "Semantic Web" can be understood from three different perspectives: vision, program and technology [6]. For this thesis, the most important view is the technology perspective. The technological view shows tools and standards which should be used within the web. For this purpose, the Semantic Web may be described by a layered architecture as represented in figure 3.6. Those layers also include the RDF, which is the first layer of the Semantic Web proper, and the RDF Schema, which is a type modelling language describing relations and classes between them [6].

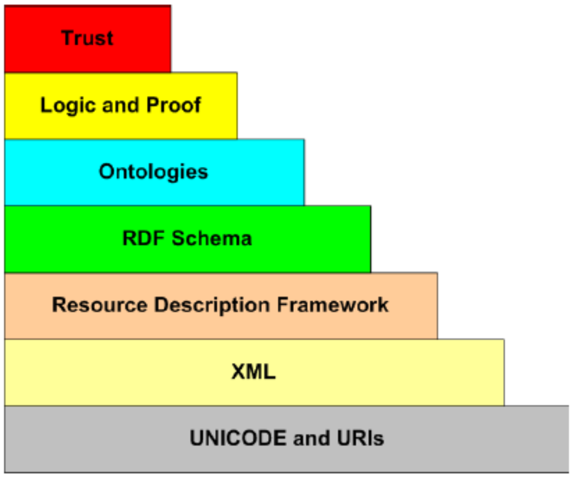

Figure 3.6.: Layers of the Semantic Web [6].

The goal of the Semantic Web is not to bring artificial intelligence to the web, but to make it machine processable. Evaluating the RDF, the main advantage is the data interoperability [50]. With data interoperability, the RDF enables easy understanding of the structure as it uses a standardized vocabulary, classes and properties [50]. Furthermore, the RDF is following a structure which helps designing an XML and prepares data for the Semantic Web [50]. Nevertheless, RDF usage has also several disadvantages. The RDF format contains constraints on the design of the realization, such as RDF within XML [51]. Moreover, the RDF uses namespaces for an unique identification of classes, resources and properties. This shows that know-how about namespaces is required. Besides the RDF format and the used namespaces, another challenge is to learn the RDF-specific vocabulary, which is used for the implementation within XML [51]. In summary, the RDF is an essential part of the Semantic Web and is widely used for web data management [52].

3.2.2. Dublin Core

The Dublin Core Metadata Element Set (DCMES) [15] contains the 15 most common elements of digital information and they were defined by Dublin Core Metadata Initiative 1995 in Dublin, Ohio. DCMES is implemented in web applications and used for indexing web pages by search engines. The motivation for DCMES is manifested in goals by [15]: Dublin Core Metadata for Resource Discovery as:

- Simplicity of creation and maintenance
- Commonly understood semantics
- Conformance to existing and emerging standards
- International scope and applicability
- Extensibility
- Interoperability among collections and indexing systems

The 15 elements of the DCMES are described in table 3.3. They group the elements in three categories: content related elements, from the intellectual property related elements and elements related to the Instantiation of the digital information.

Content	Intellectual property	Instantiation
Title	Creator	Date
Subject	Publisher	Format
Description	Contributor	Identifier
Type	Rights	Language
Source		
Relation		
Coverage		

Table 3.3.: 15 elements of the DCMES [15].

The 15 elements can be defined as follows [15]:

1. **Title** [Label: ”Title”]
   The name given to the resource, usually by the Creator or Publisher.
2. **Author or Creator** [Label: ”Creator”]
   The person or organization primarily responsible for creating the intellectual content of the resource. For example, authors in the case of written documents, artists, photographers, or illustrators in the case of visual resources.
3. **Subject and Keywords** [Label:”Subject”]
   The topic of the resource. Typically, subject will be expressed as keywords or phrases that describe the subject or content of the resource. The use of controlled vocabularies and formal classification schemes is encouraged.
4. **Description** [Label:”Description”]
   A textual description of the content of the resource, including abstracts

in the case of document-like objects or content descriptions in the case of visual resources.

5. **Publisher** [Label:"Publisher"]
   The entity responsible for making the resource available in its present form, such as a publishing house, a university department, or a corporate entity.

6. **Other Contributor** [Label:"Contributor"]
   A person or organization not specified in a Creator element who has made significant intellectual contributions to the resource but whose contribution is secondary to any person or organization specified in a Creator element (for example, editor, transcriber, and illustrator).

7. **Date** [Label:"Date"]
   A date associated with the creation or availability of the resource. Recommended best practice is defined in a profile of ISO 8601 that includes (among others) dates of the forms YYYY and YYYY-MM-DD. In this scheme, for example, the date 1994-11-05 corresponds to November 5, 1994.

8. **Resource Type** [Label:"Type"]
   The category of the resource, such as home page, novel, poem, working paper, technical report, essay, dictionary. For the sake of interoperability, Type should be selected from an enumerated list that is currently under development in the workshop series.

9. **Format** [Label:"Format"]
   The data format and, optionally, dimensions (e.g., size, duration) of the resource. The format is used to identify the software and possibly hardware that might be needed to display or operate the resource. For the sake of interoperability, the format should be selected from an enumerated list that is currently under development in the workshop series.

10. **Resource Identifier** [Label:"Identifier"]
    A string or number used to uniquely identify the resource. Examples for networked resources include URLs and URNs (when implemented). Other globally-unique identifiers, such as International Standard Book Numbers (ISBN) or other formal names are also candidates for this element.

11. **Source** [Label:"Source"]
    Information about a second resource from which the present resource is derived. While it is generally recommended that elements contain information about the present resource only, this element may contain metadata for the second resource when it is considered important for discovery of the present resource.

12. **Language** [Label:"Language"]
    The language of the intellectual content of the resource.

13. **Relation** [Label:"Relation"]
An identifier of a second resource and its relationship to the present resource. This element is used to express linkages among related resources. For the sake of interoperability, relationships should be selected from an enumerated list that is currently under development in the workshop series.

14. **Coverage** [Label:"Coverage"]
The spatial or temporal characteristics of the intellectual content of the resource. Spatial coverage refers to a physical region (e.g., celestial sector) using place names or coordinates (e.g., longitude and latitude). Temporal coverage refers to what the resource is about rather than when it was created or made available (the latter belonging in the Date element). Temporal coverage is typically specified using named time periods (e.g., neolithic) or the same date/time format as recommended for the Date element.

15. **Rights Management** [Label:"Rights"]
A rights management statement, an identifier that links to a rights management statement, or an identifier that links to a service providing information about rights management for the resource.

The main advantages of Dublin Core are the simplicity, comprehensibility and flexibility of the metadata model. It describes web resources' attributes which are commonly used [5]. A disadvantage of Dublin Core can be the missing specification of how data should be entered in the elements of the DCMES [53]. In respect to the development of the information model, the DCMES may be helpful in describing the essential elements of the GDPR Art. 20. But as a negative component, the DCMES is usually only used on online content like websites. In case of defining a DCMES-related information model, it also has to be applied to documents not available on the web. Furthermore, it would change or extend the common elements by the necessary ones for GDPR Art. 20. This may lead to a more complex information model. Moreover, the insertion type cannot be defined which might cause problems as well.

3.2.3. Evaluation of the document-based approach

Reviewing the described metadata models from the previous subsections, the document-based approach contains interesting ideas. Nevertheless, they would lead to a review of all documents, databases and their structures. Within an enterprise this means an immense workload. Thus, this approach may not be the most efficient way of assuring the fulfillment of GDPR Art. 20 for all relevant information. Another negative aspect of the document-based approach and the described metadata model is their usage, which mainly focuses on online stored data within web applications. For instance, the RDF is using URIs which are online identifiers and usually not available for offline data.

3.3. Business Object-based Approach

In contrast to the previous, this section discusses a more abstract layer for the information modelling basis - the business object-based approach. As mentioned in 3.2.3, the document-based approach includes a review and implementation on the database layer, which would influence the data structure. Thus, a high workload for huge enterprises are expected. To avoid or reduce the implementation effort, the implementation can focus on the semantic layer within a company. Here, business objects are in the spotlight.

3.3.1. Business objects according to TOGAF

Before focusing on details of a possible information model implementation, the term "business object" should be defined. One interesting definition of business object is given by the industry consortium The Open Group [7] who develop the enterprise architecture framework TOGAF 9.2 [54]. They defined business objects within their ArchiMate 2.1 modelling tool for enterprise architecture as the following:

> "A business object is defined as a passive element that has relevance from a business perspective." [7]

A more detailed explanation of the definition's meaning is given by them: A business object is representing an element which is essential in an informative or conceptive aspect. Furthermore, it is used for modelling object types within the enterprise architecture. Another important characteristic of business objects is its passive behavior towards processes (it does not trigger or perform them). With respect to the passive behavior, a business object can be accessed by different business processes, functions, interactions, events and services. Besides its accessibility, a business object can have a specialized, aggregate, associative or composite relationship with others. While a business object can be constituted by a data object or representation, its name should be a noun.

An example for a business object is modelled within the ArchiMate 2.1 Specification displayed in figure 3.7.

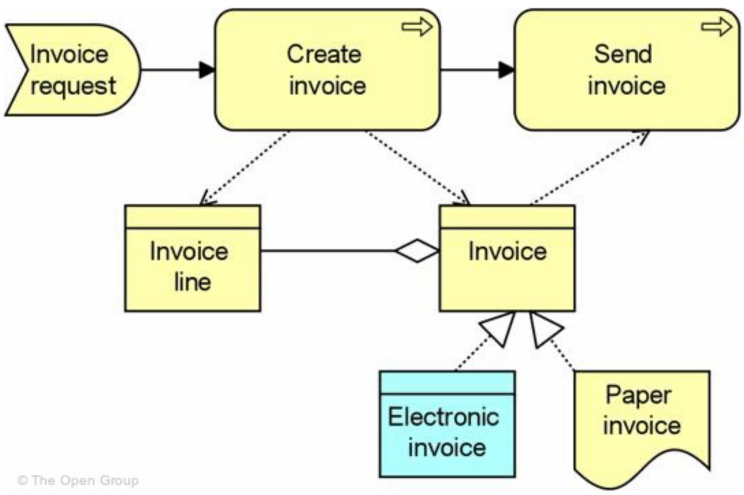

Figure 3.7.: Example of business objects according to TOGAF [7].

Figure 3.7 shows the business object "Invoice" which can either be a realization of an "Electronic invoice" or "Paper invoice". Furthermore, the business object "Invoice" is an aggregation of the business object "Invoice line". While the business process "Create invoice" is creating the business objects "Invoice line" and "Invoice", the business process "Send invoice" is accessing the business object "Invoice". This example displays the several characteristics of an business object described above by [7].

3.3.2. Business objects according to the EAM Pattern Catalog

Another definition and usage of business objects is introduced by the Enterprise Architecture Management (EAM) Pattern Catalog [55]. The Pattern Catalog describes common patterns to solve enterprise architecture issues. For this purpose, it describes management concerns, methodologies to address the concerns, viewpoints for support of methodologies and information models as basis for viewpoints [55, p.17]. Within the methodology-, viewpoint- and information-patterns, business objects also occur. Therefore, the definition of a business object according to Buckl et al. is the following:

> "An BusinessObject represents a business entity (e.g. an invoice) that is used during the execution of a business process, which performs operations (CRUD) on the Business-Object." [55, p.210]

A more detailed explanation is given by Buckl et al. [55, p.89]: The EAM Pattern gives several application landscape elements the right of usage and process of business objects. Due to parallel usage of business objects in several applications, the management of business objects is of high importance within the EAM. Furthermore, [55, p.90] explains the usage of business objects for the definition of a domain or business service. Another application of business objects is the coordination of communication between processes, services and applications.

An example for an information model is displayed in figure 3.8.

3.3.3. Business object types

For the definition of business objects within a company, the question arises whether they should distinguish viewpoints on business objects. Business objects may be of material or immaterial character [56]. For example, an order may be material on a paper, or it can be immaterial when ordering food in the restaurant verbally. Another recommendation by Hess et al. [56] is to distinguish between fast changing business objects (transaction data) and slowly changing business objects (master data). An example for master data is customer data like the name, address, date of birth, etc. as it is rarely updated by the customer. Transaction data is orders and invoices, as this data is changing and growing steadily - depending on the users' behaviour.

For the fulfillment of GDPR Art. 20, a distinction of master data and transaction data of business objects is necessary as it contains different effort of extracting the information to fulfill the requirements [29].

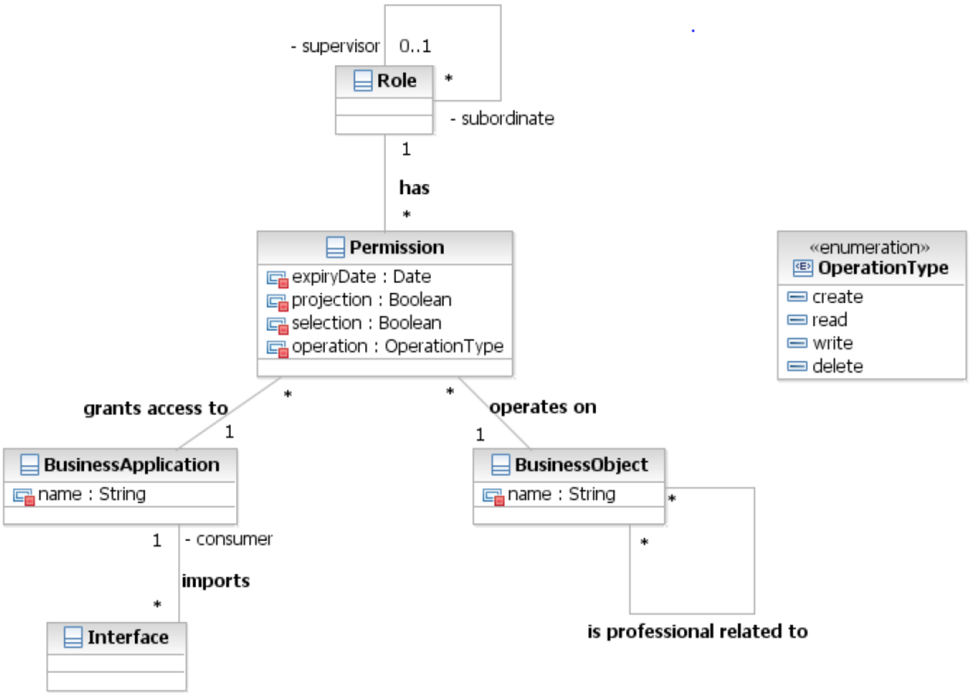

Figure 3.8.: Information model example for EAM [8].

3.3.4. Evaluation of the business object-based approach

The construct of an information model on a business object-based approach gives the advantage to show the relevant information on an abstract level. It is a more general approach and may be a kind of a reference model. Furthermore, it can focus on more relevant aspects. A disadvantage of the business object-based approach is the missing implementation details. There is no concrete explanation on how to implement this approach. As every enterprise has a specific and unique system landscape, it is very difficult to define a general approach for implementation which fits to every company. Thus, a model designed by the business object-based approach is sufficient which covers the problem domain without any additional implementation details. A deeper discussion of business object modelling is described by [57].

Part II.

Information Model Development

4. Information Model Development

After conducting the literature review, the findings should support developing an information model. This chapter will describe the procedure of information model development by starting with a discussion and justification of the chosen information model approach. Afterwards, the first draft of the information model is discussed.

4.1. Information Model Design

This chapter will discuss the approach and procedure used for the development of the information model. Therefore, the level of abstraction, the basis for the information model and the modelling language are discussed in the following sections.

4.1.1. Information model level of abstraction

To develop an information model which should help gathering required data to comply with GDPR Art. 20, several information about the model is necessary.

One important aspect which has to be discussed and aligned with stakeholders before information modelling, is the level of abstraction of the model.

The information model developed within this thesis should be independent from a company or platform used. Thus, a systematic approach for platform-independent design [9] is pursued. This approach contains several milestones [9] such as the definition of a service, the platform-independent and the platform-specific service design.

The service definition specifies the boundaries without considering the infrastructure. A service should be feasible by using a wide range of platforms which provide various interaction patterns. Applications using the service relying on the service definition should be designed at the same level of platform-independence. The upper part of figure 4.1 shows the platform-independent service definition as an example. Further, this figure displays the different levels of abstraction and provided details.

One milestone is the platform-independent service design. This phase contains the platform-independent service logic by using service components and abstract-platform definitions. Especially, the portability requirements of the characteristics of the platform have to be taken into consideration. Figure 4.2 shows the different milestones in the systematic design approach.

The last milestone is the platform-specific service design which contains platform-specific service components and a concrete platform definition. In case, the abstract platform definition does not differ from the platform chosen, the transfer from platform-independent to platform-dependent service design can be done easily.

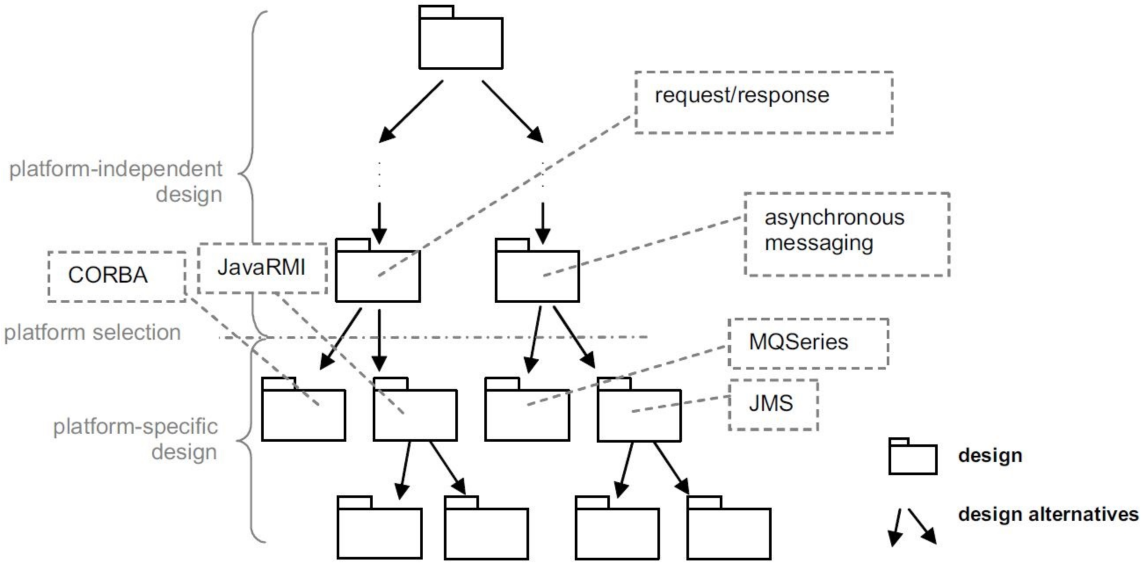

Figure 4.1.: An model driven architecture design trajectory for platform-independent design [9].

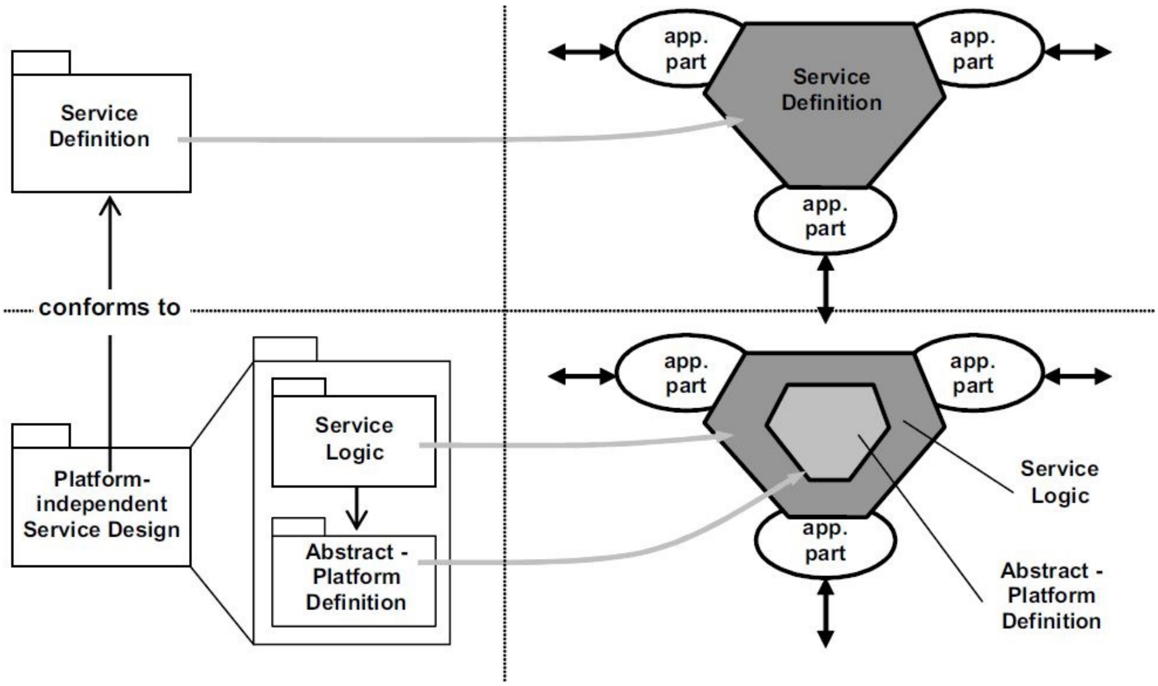

Figure 4.2.: Milestones in systematic design approach [9].

The information model developed within this thesis follows the systematic approach of platform-independent service design. In this case, the service definition contains the definition of necessary information to fulfill data portability GDPR Art. 20 and the definition of how this data can be gathered from an abstract point of view.

Besides the justification of platform-independent design, the information model should function like a reference model [3] which is described in chapter 3.1.2. This means in detail, that the developed information model should not be specific for one company but should have the ability to be transferred to several companies and industries.

In summary, the underlying information model should depict the required information to comply with data portability GDPR Art. 20 **independent** of company, industry, platform or other specific criteria.

4.1.2. Information model basis

Section 3 covers the discussion of several approaches how to develop an information model. Within this chapter, the document-based and business object-based approach are discussed.

The document-based approach considers metadata models which could be enhanced with the information requirements to comply with data portability. With regards to a company managing a high amount of documents (documents itself, data bases, files etc.), the approach of reorganizing and restructuring the existing infrastructure results in enormous work load which is not suitable for existing, large companies. For example, changing the structure of all existing invoices and its content by enhancing it with additional fields is too complicated.

Thus, the idea of the business object-based approach seems much more convenient. Within this procedure, the data, which is gathered to comply with data portability is based on business objects from a semantic view. Here, business objects can be easily linked to the required information. For example, the class “invoice” from the semantic class diagram is linked to another class which gathers the required information for compliance with data portability GDPR Art. 20. Therefore, the business object-based approach for an information model is on a more abstract semantic level which meets the challenges of platform-independent design [9].

4.1.3. Modelling language

When developing an information model, the question arises which modelling language should be used. As there are several languages available, the best fit for the use case has to be chosen.

For information modelling of a reference, platform-independent use case several modelling languages may be considered as opportunities.

According to [58], the following modelling languages are suitable for Enterprise Architecture field: *Textual Description in Natural Language, Meta Object Facility (MOF) and Unified Modelling Language (UML) Class Diagrams, Ontology Languages, Mathematical Formalization*.

The first option described by [58] is the textual description in natural language which consists of elements and relationships defined by easy natural language. Nevertheless, the construction might be mistaken or not sufficient enough for the model element descriptions.

The second language discussed by [58] MOF and UML Class diagrams are modelling languages using object-oriented concepts. UML class diagrams are commonly used and known by software developers. Regardless, the modelling language may have a deficiency in formal basis. Thus, limitations can occur in the section of where the information model may be used.

The ontology languages may be used for model fragments and relations as a more expressive power than MOF or UMLs as it is based on a formal foundation with the opportunity using automated reasoning [58]. Nevertheless, ontology languages are not commonly used in conceptual modelling.

Lastly, the mathematical formalization describes the information in formally logical or mathematical terms [58]. Though, formal or logical models are difficult to use.

In summary, according to [58] the best option for modelling of enterprise architectural problems is UML. This hypothesis is supported by [9] [59] as UML enables the representation of a service abstract and concrete platform. Thus, the modelling language used for the use case described in this thesis is UML.

4.1.4. Information modelling process

Developing an information model implies to consider several things. As described in section 3, there is a process to follow for the development of an information model. This process is described by Schütte [3] including five phases: Problem definition, model frame, model structure, model finalization and model implementation and usage.

Phase 1: Problem definition

The underlying problem for the information model developed within this thesis is to define which data is necessary to comply with data portability according to GDPR Art. 20. For the broader problem domain of data portability GDPR Art. 20, the greatest problems are the following:

1. The data controller needs to know *which* data is affected by data portability GDPR Art. 20.

2. The data controller needs to identify *where* to find the necessary data within their systems.

3. The data controller needs to know *how* to transfer the data direct either manually or electronically.

This thesis focuses on the problem of which a company may find the necessary information to comply with the legal regulation. As the second question focuses on where to find the necessary data (e.g. which data center) which cannot be displayed platform-independently, it is out of scope for this thesis. The third question covering the actual transfer of data involves several companies, which is as well out of scope for this thesis. A short introduction to data transfer is included in section 6.

Phase 2: Model frame

The model frame of the information model to be developed first should consider the specification of a business object. This business object definition may be reused from [8] as it is included in the description of the business object-based approach in section 3.3. Furthermore, it should consider the differentiation of master and transaction data which is described in chapter 3.3 as well. Moreover, the requirements identified in 2.2 should be represented by an element within the model.

Phase 3: Model structure

For the model structure the model elements described in the previous phase have to be related to each other. For model construction the UML modelling language is chosen and the first draft may be seen in the next subsection.

Phase 4: Model finalization

For the model finalization the relations of the model elements have to be designed. Before implementation and usage the model will be evaluated by conducting expert interviews to increase the quality and consistency of the model.

Phase 5: Model implementation and usage

As the developed information model within this thesis is platform-independent and a reference model for companies, an interpretation of the model is necessary for a concrete implementation and usage. Thus, an example implementation is given in section 5.3.3.

4.2. Information Model First Draft

Within this chapter, the first draft of the information model is presented. This first draft is the basis for the evaluation process in the following section.

The first draft is displayed in figure 4.3.

The information model in the first version shows the distinction of a business object in the business types ”Business Subject” and ”Business Entity”. This distinction shows the differentiation between master and transaction data.

The next part is the ”GDPR” which displays a service that is connected to the master data record of a Business Subject. This GDPR service contains the required information according to Data Portability GDPR Art. 20. Furthermore, it contains the references to other related articles within the GDPR. The GDPR service needs to gather all information required to fulfill the data portability requirements. Thus, to assure that the correct data is gathered, the GDPR service should consider respective Business Subject details like the name or ID of a person. For the application of the information model, the GDPR service has to conduct queries to request the required information from the data basis.

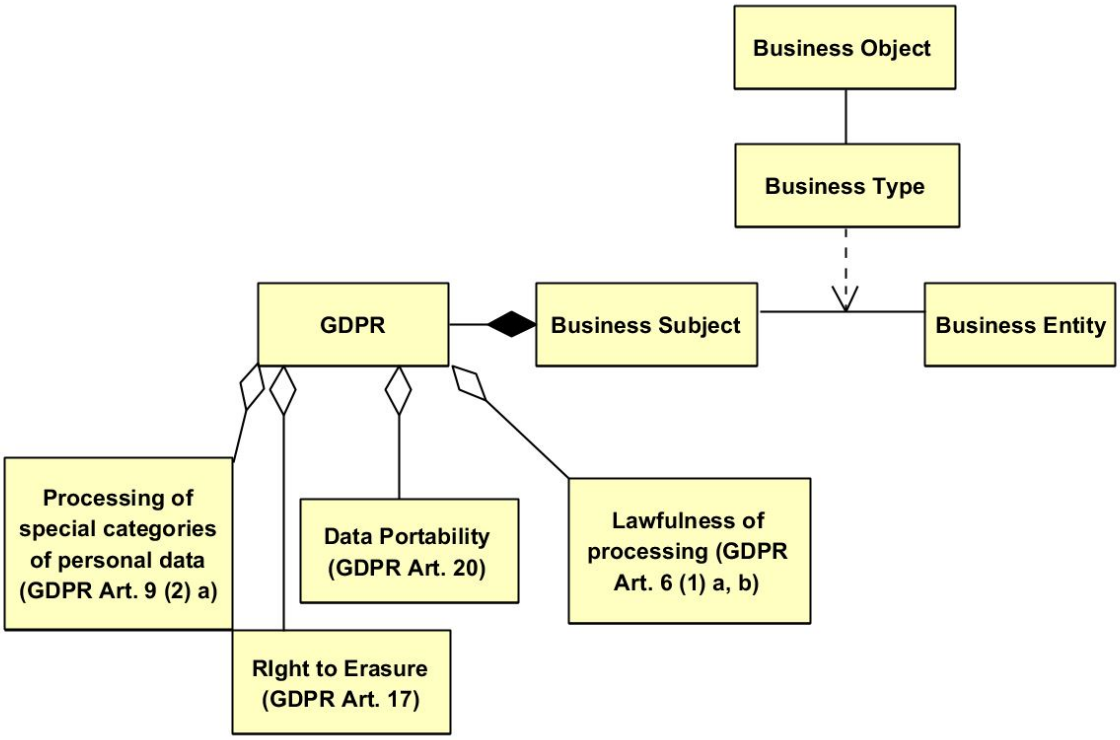

Figure 4.3.: Information model first draft.

The first draft is an abstract model which has to be adapted to the respective company and use case. It does not contain implementation details as this information model should be platform-independent. The evaluation of this model is explained and described in section 5.

5. Information Model Evaluation

Proceeding with the first draft of the information model introduced in the previous chapter, the next step is to evaluate the information model. For this purpose, an evaluation concept is presented to describe who is involved in the evaluation, how the evaluation takes place and what is the content of the evaluation. After presenting the evaluation concept, the evaluation and its influence on the information model is summarized. The outcome of this chapter is the final information model covering the data gathering to comply with data portability according to GDPR Art. 20.

5.1. Evaluation Concept

This chapter describes the concept for the evaluation of the developed information model.

For the evaluation, the information model is presented to experts. Valuable input is enhancing the information model iteratively. If the experts point out a reasonable enhancement or change to the information model, it will be adapted. This only happens if the suggestion is plausible and reasonable. After adapting the information model, the new model is used for the upcoming interview. The stakeholders for the expert interviews are enterprise architects, data management officers, data information officers and other workers in strategic information technology and compliance. The structure of an expert interview is presented as follows.

An expert interview starts with the introduction to the problem domain. Thus, the new GDPR is presented as well as the data portability according to Art. 20. Then, the resulting research goal and research questions are explained.

After explaining the research purpose, the draft of the information model is presented to the experts and discussed.

To ensure their understanding, a run through an example use case of the information model is conducted. For this purpose, an example from a social network with profile, image upload and linked images is presented as illustrated in figure 5.1.

Before asking general open questions, the information model is evaluated by Lindland's criteria [47]. These criteria discuss the semantic, syntactic and pragmatic quality of the information model. The evaluation improves the quality of the information model and helps to increase its usability.

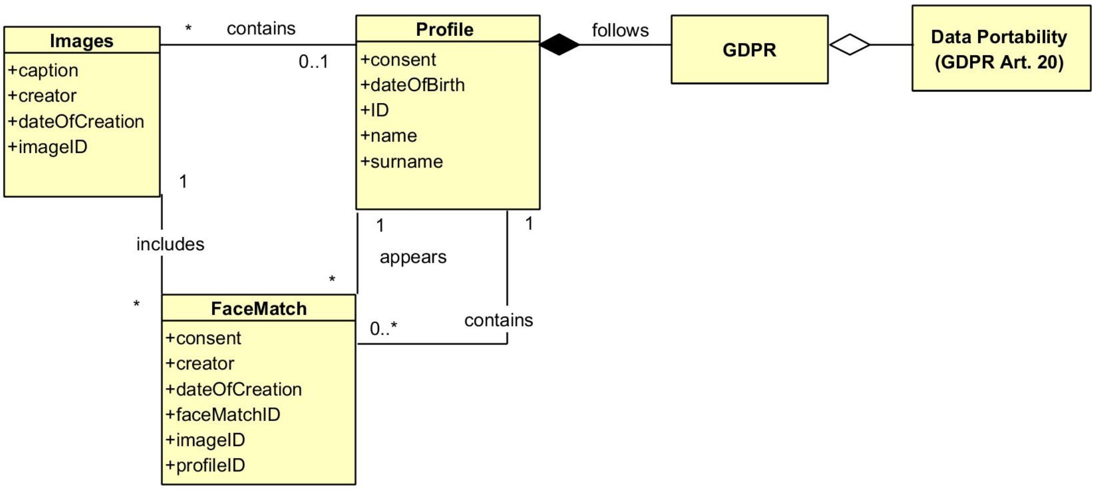

Figure 5.1.: Information model example - social network with image upload.

Finally, the expert is asked to answer the following open questions for further comments and for the improvement of the information model:

- Do you think data portability (GDPR Art. 20) is relevant for the company / customer?
- Do you think the presented information model is helpful for fulfilment of data portability (GDPR Art. 20)?
- What applications can you think of for implementing the information model?
- For which company size / branch do you think the information model may be useful?
- Can you think of an implementation of the information model? If yes, how?
- Do you have any further ideas for applications of the information model?
- Do you have any suggestions for extension of the information model?
- How are you handling the problem of data portability (GDPR Art. 20)?

During the interview, the comments and suggestions are documented in PDF comments simultaneously. This ensures the correct understanding of the experts' opinion, suggestions and key points. After each interview, the results are documented and summarized in an MS Excel sheet. Moreover, the information model is adjusted according to the reasonable enhancements or changes suggested by the experts. The updated information model is then used for the following expert interview.

After defining the evaluation concept, the evaluation process is discussed in the next section.

5.2. Evaluation Process

In summary, seven interviews for the evaluation of the information model were conducted. The following table gives an overview of the interviewees current profession, company size and corresponding industry. In interview 7, two experts were interviewed at the same time.

#	Current profession	Company size	Industry
1	Cyber Security Portfolio Manager	large enterprise	Industrial manufacturing
2	Head of IT Strategy	large enterprise	Industrial manufacturing
3	Co-Founder Compliance tool	Start-Up	IT service and consulting
4	Corporate Data Privacy Officer	large enterprise	Industrial manufacturing
5	Cyber Security Architect	large enterprise	Industrial manufacturing
6	Head of Sales for privacy management tool	SME	IT service and consulting
7	Business Intelligence (2)	large enterprise	Finance

Table 5.1.: Expert interviews overview.

5.2.1. Expert interview 1

The first interviewee is a Cyber Security Portfolio Manager with background in enterprise architecture from a large enterprise within the industrial manufacturing sector. The interviewee was responsible for innovation management in enterprise architecture for several years.

When introducing the information model example, the interviewee refers to the issue of identifying where the necessary data is stored. As a recommendation, the expert suggests an integration layer connected to a corporate master data which consolidates the necessary data gathered from different sources. When asked how a consolidation is conducted, the expert describes a procedure by first querying a request on the corporate master data which searches for data sources with the customer's ID. Afterwards, all data about the customer found on several sources is gathered and exported. Finally, the integration layer can consolidate and provide an extract for the customer.

Following the interview structure, the expert is asked to give their opinion with regards to Lindland's criteria [47]:

The syntactic quality of the model was rated by the interviewee as too general. More details are necessary for the actual implementation. Moreover, the relationship between the elements Business Object, Type, Entity and Subject was not clear enough. Another issue identified is the description of the GDPR service, which is actually a regulation not a service. The interpretation of this element is not clear enough. Furthermore, a recursion in the Business Type is suggested to cover more complex applications. With regards to the semantic quality, the expert mentions that the information model is complete and valid in relation to the problem domain. Further, the distinction between master and transaction data contributes to the validity of the model. According to the expert, the pragmatic quality of the information model is not given as the interpretation of the model is not easy to understand. They assert that as stands, the model is too complex and would improve its quality and clarity by reducing the model elements. In addition, a description of the elements should explain how to interpret them, which would provide a kind of methodology to the model.

After reviewing the quality of the information model, more general questions are asked. The expert thinks that data portability is of high relevance and that the presented information model (adjusted according to the discussed improvements) can be helpful in fulfilling the regulatory requirements. In relation to the advantages of data portability, the expert points out that even small companies are able to enter the market. Thus, a level playing field can be achieved for the whole industry. The interviewee supports that the information model is suitable for all company sizes and markets. It enables data trading within the industry, e.g. in business to business communication. However, the expert comments that data trading implies risks not everyone is skilled for. Another suggestion for further enhancement is to focus on standards in the actual data portability like the Electronic Data Interchange (EDI). Hence, the expert refers to a colleague whose interview is described in *expert interview 5*.

5.2.2. Expert interview 2

The second interviewee is the Head of IT Strategy and is leading IT Enterprise Architecture as well. The expert is working in a large enterprise within the industrial manufacturing sector.

At the beginning of the interview, the expert indicates issues regarding the fact, that it is necessary to know where the data is located in the system. Moreover, metadata about the profile is required to enable querying the data sources. Another point mentioned by the interviewee is the difficulty in regards to deleting or transferring references and backups of data with respect to data portability.

Following the review of the information model example, the criteria of Lindland are discussed. With respect to the syntactic quality, the expert suggests using an example with a more complex application to ensure that the cardinalities are set correctly. The recursive business type can be understood better by providing an example. Thus, the semantic quality of the model increases. Discussing the pragmatism, the expert evaluates the quality as good. Nevertheless, the interviewee points out that for implementation more details are necessary. For instance, the integration layer needs to know what to query on data systems and perhaps can not retrieve the data directly.

After discussing the model's quality, the expert answers several open questions regarding data portability and the information model. In general, the interviewee thinks data portability is an important article in the GDPR. Regardless, the expert comments that with the new GDPR people are forced to give consent to more processing rights to the data controller as it is rarely possible to modify the processing consent individually. Discussing the usefulness of the information model, the expert suggests reviewing the complexity of the problem as data is scattered everywhere and perhaps an analytic information model working with for example statistic methods and artificial intelligence may be more helpful. According to the expert, the complexity of data sources in some large enterprises makes the implementation of the discussed information model unmanageable. The information model can be more useful for a company with a greenfield approach. Consequently, the cost-benefit relation regarding the implementation effort of the information model and actual demands should be considered. Thus, the expert suggests using the information model for companies with a greenfield approach or companies with high demand for the enforcement of data portability.

5.2.3. Expert interview 3

This expert is a Co-Founder of a compliance start-up tool. The interviewee has experience in the telecommunication industry and founded a start-up which supports some topics of the new GDPR with the help of a tool.

When evaluating the information model according to the syntactic quality, the interviewee suggests using clearer terms than business subject and business entity. Hence, a description of the model elements should be provided. Although the expert comments that the model covers the problem of data portability, the interviewee's opinion is that it is necessary to describe the issue on a more detailed level, e.g. for one specific industry. Thus, this suggestion confirms the objective of the model as a general, platform-independent description of the problem domain. Asking the expert several open questions, the expert supports data portability, especially in industries where it is beneficial for the customer to transfer its data to another controller. Moreover, the interviewee insists on an industry-specific model to provide more details and enable specific implementation descriptions for use cases. An example industry given by the expert is the insurance sector as it provides high potential for customer changes.

Finally, the expert explains that data portability requires a standard such as in the finance or telecommunication sector. Nevertheless, developing a standard in an industry is dominated by market leaders. According to the interviewee, small companies conduct a cost-benefit relation analysis and most likely do not implement a standard for data portability as it is too high-cost. With respect to large enterprises, the expert suggests connecting the customer relationship management system with a standard Application Programming Interface (API).

5.2.4. Expert interview 4

This expert is the Corporate Data Privacy Officer of a large enterprise within the industrial manufacturing sector. Thus, the interviewee is an expert in complying and handling requirements of the GDPR.

The expert mentions the previous data privacy law in Germany which enabled data subjects to access their data prior to the GDPR. Nevertheless, after the announcement of the GDPR people have become more aware of data privacy and want to make use of the regulation's advantages. With respect to data portability, the expert mentions a draft of the European Commission which restricted data portability to companies that provide services for a information society. This restriction is not included in the final GDPR. According to the interviewee, this is imprudent as the fundamental idea of the GDPR targeted social media companies, but now it affects all industries without exception. Nevertheless, data portability is not a strongly demanded service within every industry or company. In some, there is no advantageous application for the portability of customer data, as only institutions are company's customers.

Discussing the information model, the expert criticizes the "machine-readable" format prescribed by GDPR Art. 20 as something which is not enforceable across all sectors. Concerning the syntactic and semantic quality of the information model, the interviewee suggests providing a more complex example with more than just one Business Type to improve the clarity of the model. Reviewing the pragmatic quality, the interviewee mentions a difficulty in usability of the information model as a whole due to missing details for implementation.

The expert confirms the importance of data portability, but suggests restricting the regulation to companies which provide services for information societies. Moreover, the expert often experiences customer ignorance regarding the enforcement of their rights. Furthermore, data subjects often do not consider the correct legal basis. Enforcing rights demands determination of the correct legal basis, which is not always the GDPR, but rather other already existing laws. Nevertheless, the expert understands the GDPR as a benefit because it provides a homogeneous data privacy law in Europe.

Discussing the usage of the information model, the interviewee suggests an industry-specific model for implementation, as a standard is necessary for data portability enforcement. In reflecting upon its usage in relation to the industry or company size, they think the model is not suitable for companies with a broad product range. For the implementation of data transfer, the expert recommends exploring the EDI standard (*expert interview 5*).

The expert's company is currently executing the data gathering process manually and for each individual case. A use case for data portability is rarely given as the company does not provide information society services.

5.2.5. Expert interview 5

The fifth interviewee is a Cyber Security Architect of a large enterprise within the industrial manufacturing sector. The expert worked several years as a Quality Manager in EDI operations. As two earlier interviewees referred to EDI, the knowledge about how data portability is handled by EDI is interesting in relation to the presented information model.

First, the interviewee introduces the primary goal of EDI standard: The EDI service in this company is responsible for a flawless portability of data. It is connected to the corporate master database and collects up-to-date information (e.g. suppliers) from that data source. The EDI service exchanges messages, like orders, similar to a postal service. The exchange of messages depends on the definition of the communication relation between the sender and receiver of the message (e.g. supplier and company). The sender and receiver have to agree on standards for the data portability. The sender has to define how the data is delivered to the EDI service, and the receivers need to specify how they want to receive the incoming data. In case the input format and output format of the data differ, the EDI service offers a conversion of input into desired output format. Besides format conversion, the EDI service provides fragmentation of messages or a message collection service. This means collecting messages and delivering them to the receiver at a specific time. To ensure a flawless data transfer, the EDI Service conducts a plausibility check by examining data input and data sent. For this purpose, the service examines every message.

Discussing the information model for data portability, the expert suggests implementing this model by using an additional plausibility check. This check could request the corporate master database to query for the data subject's information.

Overall, the expert evaluates the syntactic, semantic and pragmatic quality of the information model as good. Furthermore, they think the information model can be helpful for the fulfillment of GDPR Art. 20 and suggests a plausibility check for implementation. Although the data gathering procedure is currently conducted manually, the expert thinks the information model is applicable for all industries and company sizes.

5.2.6. Expert interview 6

The sixth expert is the Head of Sales for an SME company selling a privacy management tool which helps customers to comply with data privacy regulations. As the expert understands the functionality of the privacy management tool, the resulting suggestions for handling data portability are beneficial.

Before discussing the developed information model, the expert explains their company's procedure for dealing with data portability. The tool, provided by the company uses as basis the "records of processing activities" GDPR Art. 30. These enable identification of storage location of the data needed. According to the expert, data portability requests are usually extended to 90 days as the execution is complicated.

Reviewing the usage of the information model, the interviewee mentions that their company had already tried something similar. They intended to implement a direct connection to the server for querying the necessary information, but it did not work out as it was too complex due to the system structure. Moreover, they agreed that the actual querying is platform specific and cannot be generalized. Thus, the expert recommends a kind of workflow guidance which describes how to gather the data. According to the expert, the company's tool can scan the record of processing activities and filter out the data relevant for complying with GDPR Art. 20.

Discussing the syntactic quality, the expert criticizes the use of the name "GDPR" for the service which gathers the required information. As the information model focuses on a set of requirements of the GDPR, it should have a name including "service" or "tool". A review of the semantic quality evaluates the distinction between master and transaction data as very good. Nevertheless, the expert recommends a further distinction between original data from the data subject and processed data from the data controller. Those should be included to improve the completeness of the model. Talking about the pragmatic quality of the model, the expert assesses it as good since there is a transferability of the information model to the company's existing tool.

However, the expert mentions the issue that automated portability is not in the interest of the companies. Thus, they tend to invest minimal effort in complying with GDPR Art. 20. The expert thinks the information model is helpful for complying with GDPR Art. 20 and is suitable for every industry and company size. For enhancement, the interviewee suggests representing what elements of the information model are internally and which are externally given. For instance, the requirements of GDPR Art. 20 are externally given by the regulation, while data of a Business Subject or Business Entity is managed internally. Another recommendation for enhancement is to include personal consulting, which helps to identify and formulate company specific task management. As workers who need to fulfill the requests for data portability are unfamiliar with GDPR requirements, they need task descriptions for the correct execution.

5.2.7. Expert interview 7

The last interview was conducted with two experts who are working in the Business Intelligence department of a large enterprise within the finance sector. One of the experts has years of experience in technical data management. The other expert was a consultant who supported in establishing the Business Intelligence department.

After explaining the idea of the information model, the first issue discussed is the data querying. According to the experts, for most queries the Data Ware House (DWH) needs to be contacted, similar to customer master database. Talking about the data flow mapping in companies, the interviewees think querying on basis of data flows are possible, but only in small companies or companies of information societies (e.g. Google, Amazon, Facebook etc.). For companies that exist since decades in industry it is very difficult to assure data flow mapping, especially when more than one DWH exists. Besides the data flow difficulty, the experts point out the problem of data gathering within the integration layer which may not have access to all data sources.

Discussing the example of the information model, the interviewees think it can be applied by setting up queries using the DWH. Nevertheless, they emphasize that the difficulty increases with the number of DWHes. Moreover, they mention to consider another legal basis than the GDPR in fulfilling requests of the data subject.

Next, the experts give feedback regarding the syntactic, semantic and pragmatic quality. They suggest reducing the model elements by deleting the Business Type and Business Object as they make the model too complex and do not add any value. Moreover, they recommend linking the elements Business Subject and Business Entity, as a Business Entity does not exist without a Business Subject. Regardless, the experts think the model displays the problem completely and is transferable to a specific application.

Talking about the information model application, the interviewees recommend it for a greenfield approach or for companies of the information societies, whose business is trading with customers' data. Furthermore, the experts think it is useful as a starting point for a company that wants to implement data portability. To improve the information model's quality, they suggest reducing the model elements as described before which enables an easier general application.

In general, the experts think the data subject's focus is more about which data is stored about them, not the actual data portability. Moreover, they point out that the data portability itself is an easier topic than the identification of data a company is storing about the customer.

5.3. Final Information Model

This section summarizes the recommended changes to improve the information model by the interviewees. The basis for the changes are suggestions from experts identified in interviews. Further, the experts comments regarding the GDPR in general are aggregated. Finally, an explanation of the information model and its elements and an example application on a Twitter use case is given.

5.3.1. Interview suggestions influencing the information model

Some improvements mentioned by interviewee 1 are to introduce an integration layer, provide a description of how to interpret the model elements and define the relationship between Business Object, Type, Subject and Entity correctly. Moreover, they mention a recursion of the Business Type. Thus, the information model is enhanced with the element's descriptions and a review of the relationships between the Business elements. Another recommendation by interviewee 1 and 6 is to change the name of the "GDPR" element as it is an unclear wording for a service description. Thus, the name of the integration layer is changed into "GDPR Data Gathering Service". According to expert 2 and 4, the recursion of the Business Type should be displayed within a more complex example. Expert 3 mentions that an undefined usage of Business Entity and Subject is causing confusion. Therefore, a more precise description is added for those elements. Expert 1 and 7 initiate to reduce the model elements. Expert 7 recommends reducing complexity by deleting the Business Object and Type element as it does not provide additional value or meaning to the information model. Moreover, they suggest connecting the Business Entity with Business Subject because an entity would not exist without a subject. The model does not consider the recommendation of interviewee 6 to add a workflow guidance of how to collect the required data, as it needs platform-specific information. Furthermore, expert 6 suggests distinguishing between original and processed data. The filtering within the GDPR Data Gathering Service should consider this distinction. The purpose of this information model is to show a general, platform-independent model. The experts comment with respect to detail level are confirming the model's objectives.

After applying the suggestions the final information model is displayed in figure 5.2.

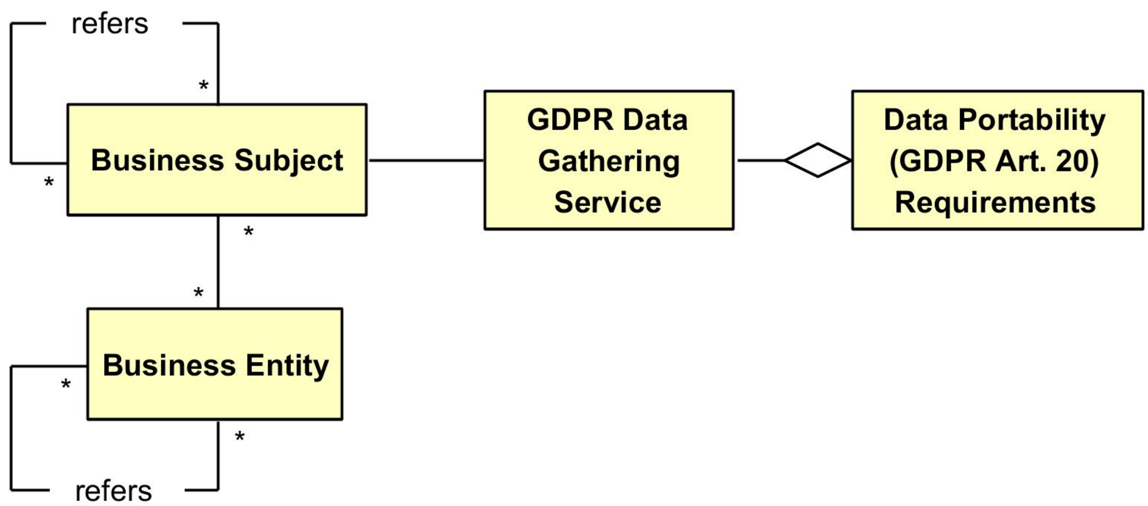

Figure 5.2.: Final information model.

Along with the information model a description of the elements is provided:

- **Business Subject:** A Business Subject is a type of business object which is stored as master data. Examples for a Business Subject are product or customer records. For this thesis, Business Subjects which are directly related to a person, such as a customer's record, are in focus.
- **Business Entity:** A Business Entity is a type of business object which is created by and directly connected to a Business Subject. In other words transaction data (e.g. invoices, orderes).
- **GDPR Data Gathering Service:** This service is an integration layer which requests data sources of the required information for the data transfer.
- **Data Portability (GDPR Art. 20) Requirements:** This class provides a list of information which are necessary to fulfill data portability. It is the basis for the GDPR Data Gathering Service, as it provides the properties which the service queries. Those requirements have been identified in section 2.2.

5.3.2. Additional interview outcomes regarding the GDPR in general

In contrast to the previous subsection, this part summarizes the interviewees remarks regarding the GDPR and data portability from a general perspective.

As most interviewees see the relevance of data portability, expert 1 emphasizes the market entry opportunity for small companies which supports the level playing field. Further, they refer to the new possibilities of data trading (e.g. B2B communication) but considering implicit risks as not everyone is skilled enough for responsible handling. With respect to data portability, the interviewee 2 points out the issue with non-existing information about how to handle references and backups and whether those are affected as well. According to experts 7, the data stored about a data subject is of higher interest than the actual transfer of this data. Hence, the data gathering process is more difficult than data portability itself.

For the data portability implementation, expert 2 and 3 think the companies should conduct a cost-relation analysis to identify whether they will invest a lot of money for an automated data portability process. Further, interviewees 2, 3 and 7 suggest to use the information model for a greenfield approach and for data-driven companies, as those know where the data is located. Expert 6 refers to Art. 30 of the GDPR "Records of processing activities" as the basis for data portability enforcement. By documenting which data is processed and where it is located, a company can identify which data is the necessary for complying with GDPR Art. 20 and where this data is located. Moreover, the expert 6 comments that the data portability requests are usually extended to 90 days as it is very difficult to fulfill the requirements. Further, they mention that companies invest minimal effort to enable automated data portability (GDPR Art. 20) as it is not in their interest to loose customers.

Regarding the development of a data portability standard, expert 3 recommends to focus on an industry where customers profit from the transfer of data. Though, the development of standards would be influenced by market leaders, which is mentioned negatively by expert 3 as well.

Disadvantages regarding the GDPR in general has been mentioned by expert 2, 4 and 7. Expert 2 thinks the companies want to profit from the GDPR by claiming more consent to processing rights as people often cannot individually select their consent agreement. Expert 4 refers to the previous existing German data privacy law which enabled data subjects already with several rights, just like the new GDPR. Through rising awareness people now start enforcing their rights, which leads to an issue experienced by interviewee 4, without knowing the correct requirements for right enforcement. In general, experts 4 and 7 criticize that the GDPR concerns everyone, even though it was restricted to companies providing services for information societies in the draft versions. According to them, some companies do not provide any information society services and thus, the requirements to comply with the GDPR are not applicable, such as for sport clubs. Moreover, the definition of the correct legal basis is of high importance for the interviewees 4 and 7 as the GDPR is not always the right basis. For example, the right of using pictures with other people is controlled by the copyright law. This already handles situations with pictures. Thus, the GDPR will not be the legal basis if someone wants to claim some rights on pictures.

5.3.3. Example application of the information model

Within this subsection, an example is provided to show how the information model can be used.

Starting with the provided general information model, the business objects within a company have to be identified first. Thus, the data model of the company should be discussed.

The social network *Twitter* and its data model [10] can be used as an example as the main objective of the GDPR are data-driven companies. The data model is displayed as UML class diagram in figure 5.3.

[10] explains that the users are Twitters primary class. A user is able to create Tweets, send Messages to other Users, is following other Users and can be grouped in Lists.

Within this class diagram the *User* class is the Business Subject, while the others (Tweet, Message, List, UserTag and HashTag) are Business Entities. Adapting further ideas of the developed information model, the GDPR Data Gathering Service is connected to the User class. Finally, the example enhanced with the information model's service is displayed in figure 5.4.

After the enhancement of the data model, the GDPR Data Gathering Service should now query the database. The required information which need to be known for data portability are identified in section 2.2. These should be queried on the database:

- personal data concerning the data subject
- consent pursuant : DATE
- specific categories of personal data : BOOLEAN
- explicit consent pursuant : DATE
- contract pursuant : DATE
- erasure date : DATE

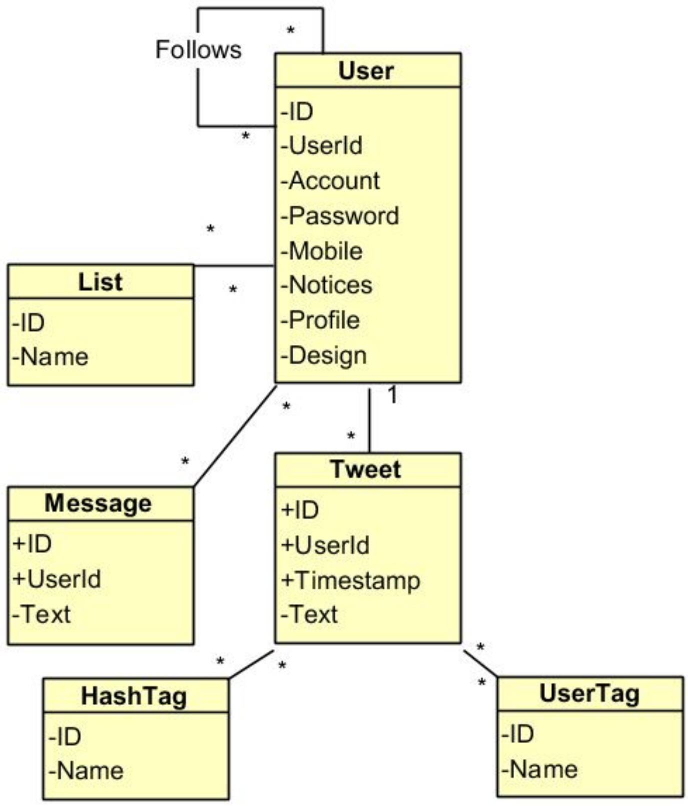

Figure 5.3.: Twitter class diagram [10]

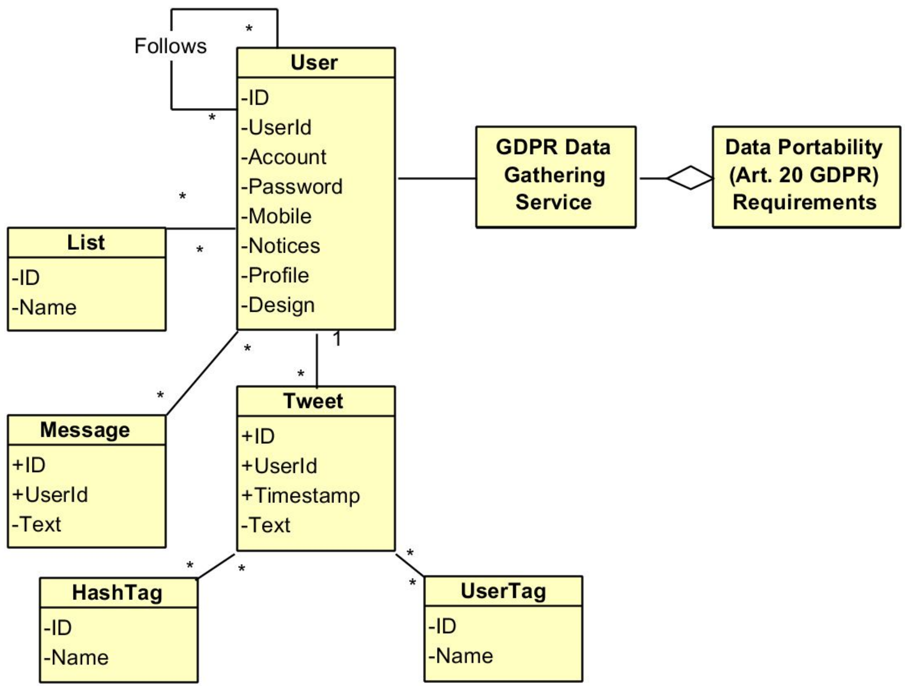

Figure 5.4.: Twitter class diagram enhanced with GDPR Data Gathering Service.

Within this application, the users personal data has to be identified by filtering for the user's ID and querying the user and its related entities.

For the Twitter example, the assumption that the data source is a SQL database is made. For this case, an example query to collect all Tweets related to the User with the ID 12345 is:

```
SELECT *
FROM User
INNER JOIN Tweet ON User.UserId = Tweet.UserId
WHERE UserId = 12345;
```

The result returns a list of the users profile and its corresponding Tweets. Those information are personal data concerning the data subject and provided by the data subject, as the data subject publishes the Tweets by themselves. The Tweets are transaction data and posting them on Twitter gives Twitter the consent to publish them. Thus, the *Tweet.Timestamp* gives information about the consent pursuant. As this data is not a specific category of personal data, it is not necessary to provide an explicit consent pursuant. The profiles contract pursuant and erasure date must be stored in the User class. The assumption is that those information is stored separately in internal databases. As there is no further information published about the internal Twitter data structure, this information query is not specified. Additionally to this query, the tweets which refer to the user are of interest as well. These tweets are data of third persons as it is created by others. Nevertheless, if the user's name is tagged within another user's tweet, then this information is of relevance as well.

6. Information Model Enhancements

After the development and evaluation of the information model, the question arises which steps have to be taken next. Therefore, this section summarizes several options how to proceed after the creation of the information model.

The information model displays the extraction of necessary information from data sources for complying with data portability GDPR Art. 20. The enhancement recommendations focus on three different viewpoints: First, a deeper discussion of the information model with more details for implementation. Second, the enhancement of the information model with requirements from other articles of the GDPR. Third, a further discussion of the actual transfer of data from one controller to another.

6.1. Deeper Discussion of the Underlying Information Model

The perspective of the current information model is very high and does not provide any implementation details. As shown in the interviews, many experts recommend an industry-specific information model. This means, a more concrete description of the model elements should be given. Referring to the milestones of the systematic design approach [9], the current information model provides the platform-independent service design. To enrich the model with implementation details, the model needs interpretation into a platform-dependent service design. Thus, industry-specific descriptions, tools and platforms need to be identified. An implementation guideline can be set up for a specific industry. It is similar to the development of a technical standard.

6.2. Consideration of further Articles of the GDPR

As it stands, the information model displays the data gathering procedure to comply with GDPR Art. 20. As the GDPR has many different articles, only data portability and its references are described in the requirements analysis. For further enhancement, it may be worth analyzing other articles of the GDPR as well. Thus, other requirements are identified and the "GDPR Data Gathering Service" can be enhanced by querying and consolidating additional information.

6.3. Concepts enabling Data Portability

Besides enhancing the information model with detailed implementation descriptions or consideration of other articles, it is interesting to discuss the actual transfer of data and its technologies, as well. Thus, several ideas such as the minimal effort, the EDI, DTP and Solid are discussed in the following subsections.

6.3.1. Article 29 Working Party guideline on data portability

The Article 29 Working Party published a guideline on the interpretation of data portability requirements [30]. They discuss various interpretation issues of the GDPR. Further, they provide information about how portable data must be provided at least. According to [30], data controllers should provide a dataset in formats like XML, JSON or CSV if there are no formats commonly used within the industry. Additionally, metadata information for an interpretation of the dataset is required. The inclusion of metadata on a high level needs to consider the protection of trade and processing secrets of the old data controller. Nevertheless, the new data controller should be able to interpret and understand the provided dataset with metadata information. Thus, it does not hamper the data subject's re-use of the data.

Besides the proposal for minimal effort format, [30] refers to the European Interoperability Framework, which encourages organizations to participate in their approach of interoperability of public services. Even though Article 29 Working Party recommends a format for industries without common standards, the technical feasibility of data portability is assessed on case by case basis.

6.3.2. Data portability with EDI

Companies feel the urge to transfer data internally and externally. EDI has been developed in 1974 in the U.S. and is covering the electronic data exchange of business transactions [60]. This business transactions can happen either internally in a company or externally between companies. For instance, a business transaction can either be an invoice or orders.

The task of EDI is to serve as "electronic post" by receiving and forwarding transactions [60]. The advantage of using EDI are additional services, like the converting of incoming transactions into another format which is forwarded to the receiver. The format is specified in a communication contract, which contains all necessary information like data structures [60]. The format conversion is an advantage for direct electronic processing of the exchanged data in the company's systems [60]. For data transfer between companies, a specific standard has been developed for EDI - the Electronic Data Interchange For Administration, Commerce and Transport (EDIFACT) [60]. Nevertheless, EDI supports several other transfer formats like CSV, JSON etc. For the usage of the EDI service, four components are obligatory [11]:

- Communcation: Data Portability and Transfer Protocols
- Data Formats and Message Standard
- ERP System Software
- Own EDI Software / Infrastructure or EDI-Clearing

Figure 6.1 displays an example of an EDI Service.

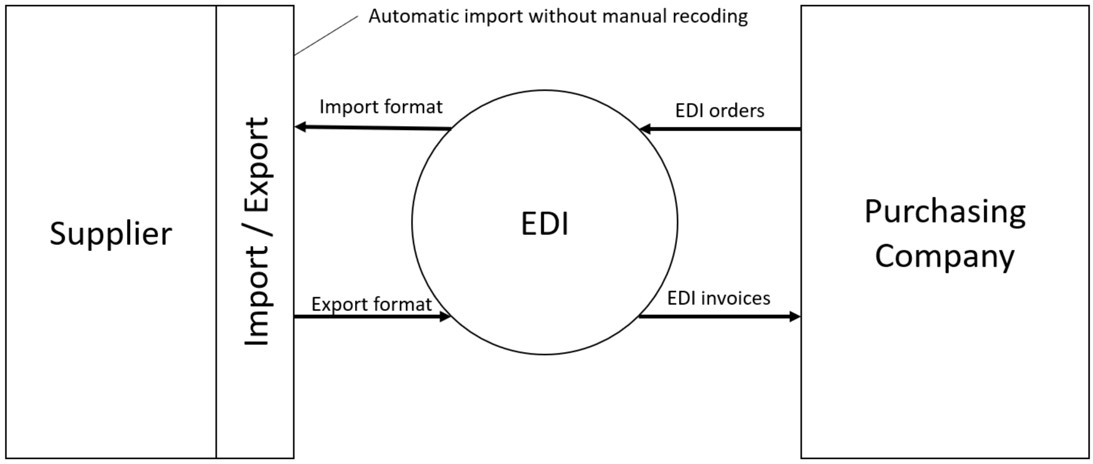

Figure 6.1.: EDI example - automatic import and export. [11]

Examining EDI usage in current time, it is still a widely used standard for digitalizing B2B business processes [61]. With respect to data privacy management of the data transferred via EDI, there are no current information available. Nevertheless, EDI is an often used standard for the automatic transfer of data internal and external of companies. For the use of EDI, receiving and sending messages, a communication contract is necessary, it is not really scalable for the data portability realization. It means that data cannot be transferred to an unknown company without a communication contract. Regardless, it can be a first option if an EDI service is already in use within the company.

6.3.3. Data Transfer Project

A major project is the Data Transfer Project (DTP) [62] which works on Data Portability and is a cooperation of Google, Facebook, Microsoft and Twitter. According to their website [62], the project was founded in 2017 as an open-source project. The main objective is to develop a service-to-service data portability platform as a part of Googles' preparations to meet the new GDPR requirements [63]. Before initiating this project, Google already set up a service called "Google Takeout" in 2011 which enabled registered users to export personal data from more than 30 Google services [64]. Meanwhile, the service is renamed into "Download Your Data" providing the data which is stored on several Google services. In consequence of GDPR Art. 20, Google wants to expand their tool by enabling direct transfer between similar service offerings like Facebook, Microsoft or Twitter [63] [65]. The whitepaper published by the DTP in 2018 [12] summarizes the background and points out the future of a service-to-service portability for data. Moreover, it contains information about the project's architecture as well as system components.

According to the DTP [12], without a centralized project each provider would have to maintain and to build adapters for API's and data formats. With the help of the DTP, the workload of maintaining multiple adapters and API's can be reduced by the API that supports the DTP Data Models. Moreover, [12] refers to standard formats when talking about DTP Data Models. The DTP solution is displayed as architectural picture in figure 6.2

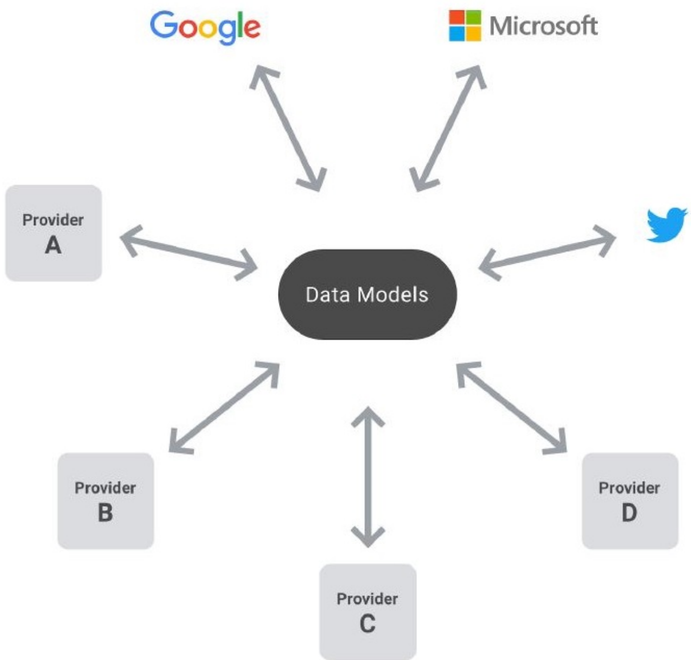

Figure 6.2.: Communication channels with the DTP. [12]

The main idea of the DTP is to use adapters that provide a translation of a provider's API to a DTP Data Model and vice versa [12]. Besides the transfer of API's to DTP Data Models, the DTP contains authentication adapters for authenticating the providers identity [12]. The DTP is a project dominated by market leaders with the objective to develop a standard service-to-service portability [12]. So far, companies like Facebook, Google, Microsoft and Twitter are officially contributing to the project. According to the website, everyone who is interested in using this new procedure can contribute and participate since the project is open-source [62].

In relation to data portability, the DTP project determines data models for business objects which are commonly used in social networks (e.g. pictures, profile data etc.). These data models are standardized for the DTP. Participating companies now have to adapt or convert their internal business objects to the DTP data models to enable data portability with it. This standard implies advantages for the further enhancement or usage of the developed information model within this thesis. Thus, implementation details for the social media sector can be given, as (under assumption) all companies would use the same data model and structure. In that case, the data gathering process can be explained easier and with more details.

6.3.4. Solid platform

Mansour et al. [13] published a concept of a decentralized platform for applications of the Social Web called Social Linked Data (Solid). The objective of this platform is to enable users more control over their data in social networks. The user's data is stored in Web-accessible personal online datastores (or pods). The Solid platform provides mechanisms for authentication and authorization for the pods. The user configures the access of applications to the pods and can switch between applications anytime. Moreover, data stored in pods can be reused and transferred between applications. The idea of enabling easy data portability between applications is matching the required objective of interoperability according to GDPR Art. 20. Figure 6.3 displays an example architecture of the Solid platform.

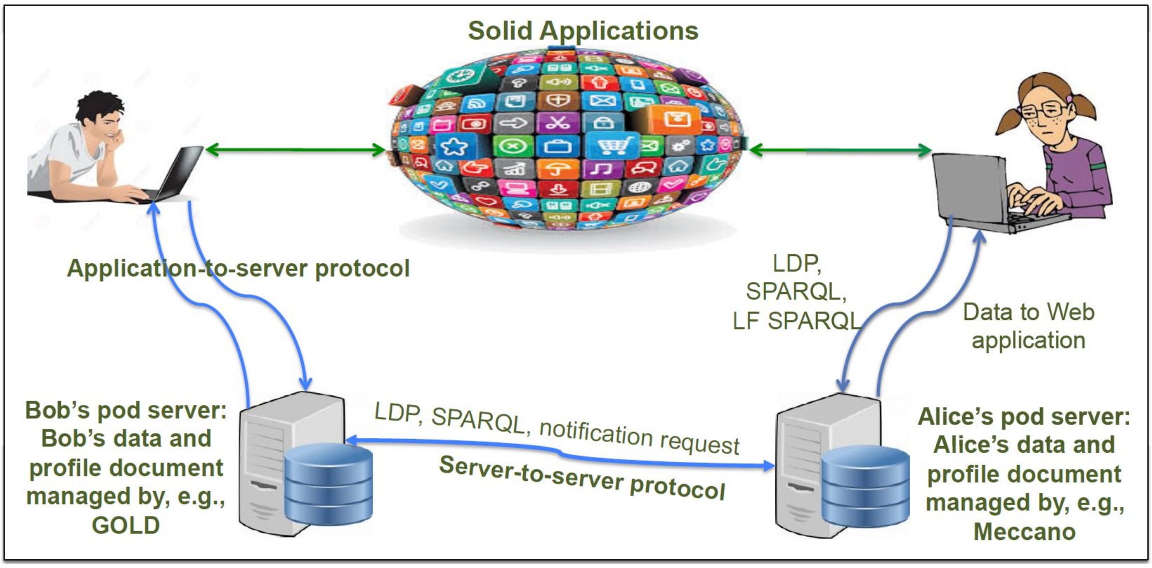

Figure 6.3.: Architecture of Solid platform [13].

The innovative idea of [13], enables user's to control their data and its usage completely. According to this architecture, the companies would not have to gather the data to comply with data portability according to the GDPR Art. 20 as the data is managed by the data subject themselves.

7. Conclusion

This thesis makes data portability according to the GDPR Art. 20 a subject of discussion. As introduced in section 1, the customer's awareness of data privacy is increasing. By enforcement the GDPR, data privacy laws in Europe are harmonized and everyone can take advantage of the new possibilities provided by the regulation.

This section discusses the main outcomes of the thesis. The research goal is developing an information model, which supports companies in collecting data which is required for complying with the GDPR Art. 20. For this purpose, section 2 discusses data portability in general and conducts a requirement analysis to identify the necessary information for compliance with data portability GDPR Art. 20. As a procedure for the identification of the required information, a systematic approach by Breaux [22] has been conducted. This uses goal mining from policies and derivation of semantic models by semantic parameterization to identify the required activities for compliance. The semantic models were used for a discussion of which information is necessary for complying with data portability. The following seven points are the identified information for compliance:

- personal data concerning the data subject
- consent pursuant
- explicit consent pursuant
- specific categories of personal data
- contract pursuant
- erasure date
- public interest

Section 3 describes the results of the literature review according to information models. This includes the basic term definitions of "model" and "information model". Furthermore, a process for information modelling is described by five phases: Problem definition, model frame construction, model structure construction, model finalization and model usage [3, p.75]. This process has been the basis for the development of the information model. After information modelling, the evaluation of such models is required. Hence, two evaluation methods [3] [47] were compared. Finally, the evaluation method based on quality aspects (semantic, syntactic and pragmatic) from Lindland [47] outweighs the model principles of [3] due to existing applications and good reputation of the evaluation method of [4]. Section 3 concludes with a comparison of a document-based and business object-based information model. The document-based approach discusses metadata models like the Dublin Core as well as the RDF in relation to the Semantic Web. The document-based approach has been evaluated as not suitable as it would affect document and database structures and therefore implies high workload and costs for companies. First, the business object-based approach defines "business objects". Next, the applicability of a business object-based information model is part of this comparison. As this approach can be applied on a semantic level, the implementation and application is easier. Moreover, a differentiation of business object types into master and transaction data [56] can be done. As the effort for gathering master data is less than that of transaction data, a distinction is required and recommended. Thus, this is considered for the information model development.

As defined in the first two research questions, the literature review of information models and the requirement analysis of data portability GDPR Art. 20 should provide a basis for the information model development and evaluation. The learnings of previous sections helped to develop the first draft of an information model. Following the information modelling process described in section 3, several aspects and outcomes for the information model were identified. Hence, the model's abstraction level was assessed. As the objective of the information model should be independent of company size or industry, a high abstraction level such as a reference model is chosen. This hypothesis is supported by the platform independent development of a service [9] as well. Next, the document-based and business object-based methods for an information model were compared. For the underlying use case, the business object-based approach has been evaluated as the best-fit due to the semantic description level and less workload and costs for companies. Further, multiple possibilities of modelling languages have been discussed. The language UML has been chosen for the information modelling as it is recommended by several sources [58] [59] within the area of enterprise architecture. Finally, the information modelling process of [3] have been finalized and the first draft of the information model has been designed, including an explanation of the components.

Section 5 continues with the evaluation of the information model draft from the latter section. As an evaluation method, expert interviews were conducted. To structure the interviews an evaluation concept shows the interview's content. The concept has been applied to all seven expert interviews. The structure of every interview included the presentation of the interviewee, the comments on the information model, an evaluation of its quality criteria (Lindland [47]) and general suggestions and opinions on data portability and the GDPR. After each interview reasonable suggestions were applied to the information model. Thus, it increased its quality iteratively. In the end, the final information model (figure 5.2) with a summary of the changes and explanation of the model's elements is described in section 5. Moreover, the interview's outcome included additional interesting comments regarding the GDPR and data portability, which are not related to the information model itself. These have been summarized as well in this section. Further, an application example on Twitter's class diagram [10] was given. To assure better understanding and interpretation of the information model, it was applied to a use case and some SQL statements for gathering the required information were provided.

The last section of this thesis discussed several enhancement possibilities. The first idea is to enhance the existing information model with more implementation details for one specific industry. This has been one essential outcome of the expert interviews as it was criticized often. The second proposal is the consideration of other articles of the GDPR. Here, the information model can be enhanced by complying with other requirements of the GDPR. The third possibility discusses several concepts which enable the actual data portability. The easiest implementation is the minimal required effort described by Article 29 Working Party Guideline [30]. Here, an export of the personal data in XML, JSON or CSV format is enough. Another possibility is the EDI service, which enables the direct transfer of data between companies [60]. This procedure is used for the electronic and automatic exchange of data between supplier and customer. For enabling EDI, the communication contract between two parties needs to be specified. This is an administrative task which implies high effort if the industry is broad and there is no standard for data portability. Hence, its usability needs to be discussed. Furthermore, the DTP project from Google, Facebook, Twitter and Microsoft is described in section 6. In detail, the DTP uses adapters that provide a translation of a provider's API to a DTP data model and vice versa [12]. They are trying to establish a standard for data models in social media to comply with GDPR Art.20 and other regulations. This can be seen critically, as the market leaders can influence the standard definition and therefore impede adaption for smaller companies. Lastly, the concept of Solid [13] describes how people can manage their data on their own and allow applications and others the access to their own data pods. This concept displays data portability enforced by the users' decision of who has access to their data. With this approach, the relevance of data portability needs to be reviewed as the customer can transfer their data independently.

As discussed within this section, there are several opportunities for enhancing the developed information model. But, also negative sides have to be considered, such as the applicability of data portability in specific industries where companies are not data-driven. Regardless, the data privacy awareness will increase in the future and laws like the GDPR will become more important for data handling. Further, the GDPR rarely contains concrete explanations. Thus, the jurisdictions regarding the GDPR will be interesting as those will show the scope for interpretation.

Appendix

A. Goal Mining Results of Referenced Articles by Data Portability GDPR Art. 20

This gives an overview of the goal mining results of Article 6 (1) a, b; Article 9 (1) (2) a and Article 17 of the GDPR.
The goal mining of GDPR Art. 6 **"Lawfulness of processing"** result in the identified goals described in table A.1.

Actor	Action Word	Subject Type	Conditions
DC	PROCESS	personal data	if...
DS	PROVIDE	consent of personal data (for specific purposes)	to processing by DC
DC	REQUIRE	processing of personal data	to fulfill contract
DC	REQUIRE	processing of personal data	for DS to be able to enter a contract

Table A.1.: Goal mining of GDPR Art. 6 (1) a,b.

Another article referred in by data portability (GDPR Art. 20) is article 9 paragraph 1 and 2 a which is discussing the topic of **"Processing of special categories of personal data"**. The goal mining process identifies the goals, described in table A.2.

Actor	Action Word	Subject Type	Conditions
DC	PROHIBIT	processing of personal data	of racial, ethic origin, political opinions, religious, philosophical beliefs, trade union membership, genetic biometric data, health, sexual orientation
DS	GIVES	explicit consent to processing of special categories of personal data	to DC

Table A.2.: Goal mining of GDPR Art. 9 (1) (2) a.

Finally, the last article referred by data portability is article 17 **"Right to erasure (right to be forgotten)"**. The result of the goal mining process is displayed in table A.3.

Actor	Action Word	Subject Type	Conditions
DC	ERASE	personal data	concerning DS
DS	OBTAIN	personal data deletion	from DC
DC	ERASE	personal data	without undue delay when personal data no longer necessary in relation to purpose it was collected
DS	WITHDRAWS	consent	to processing by DC
DS	OBJECTS	processing pursuant	to DC
DC	ERASE	personal data	when unlawfully processed
DC	ERASE	personal data	for compliance with legal obligation in Union or Member State law to which controller is subject
DC	ERASE	personal data	collected in relation to information society services
DC	INFORM	personal data to erase	to other controllers processing this personal data
DC	INFORM	personal data to erase	taking account of available technology, cost of implementation
DC	INFORM	personal data to erase	if there are any links, copies, replications of those personal data
DC	NOT ERASE	personal data	if..
DC	REQUIRE	personal data	to processing for exercising the right of freedom of expression and information
DC	REQUIRE	personal data	for processing by Union or Member state law to which controller is subject
DC	REQUIRE	personal data	to perform task carried out in the public interest
DC	REQUIRE	personal data	in exercise of official authority vested in controller
DC	REQUIRE	personal data	for public interest in area of public health
DC	REQUIRE	personal data	for archiving purposes in public interest, scientific, historical research purposes, statistical purposes
DC	REQUIRE	personal data	if right to erasure is likely to render impossible or seriously impair the achievement of the objectives of that processing
DC	REQUIRE	personal data	to establish, exercise or defend legal claims

Table A.3.: Goal mining of GDPR Art. 17.

B. RNLS of Referenced Articles by Data Portability GDPR Art. 20

The referenced article 6 (1) a and b contributes the RNLS#14 until RNLS#17 from table B.1.

RNLS#	Ref#	Statement
14		DC process personal data.
15	14	DS provide consent to RNLS#14 for one or more purposes.
16	14	DC RNLS#14 to fulfill contract with DS.
17	14	DC RNLS#14 to enable DS to enter contract.

Table B.1.: RNLS of GDPR Art. 6 (1) a,b.

The RNLS for GDPR Art. 9 (1) (2) a are defined in table B.2.

RNLS#	Ref#	Statement
18		DC process personal data of racial, ethic origin, political opinions, religious, philosophical beliefs, trade union membership, genetic, biometric data, health, sexual orientation.
19	18, 20	DC requires RNLS#20 to RNLS#18.
20	18	DS gives explicit consent to RNLS#18.

Table B.2.: RNLS of GDPR Art. 9 (1) (2) a.

The table B.3 shows the RNLS derived from article 17 which is referred by RNLS#11 and RNLS#12 from article 20 of the GDPR.

RNLS#	Ref#	Statement
21		DC erase personal data concerning DS.
22	21	DS request data deletion to RNLS#21.
23	21	DC RNLS#21 when no longer necessary in relation to the purpose it was collected.
24	21	DS withdraws consent to processing to RNLS#21.
25	21	DS objects processing pursuant to RNLS#21.
26	21	DC RNLS#21 when unlawfully processed personal data.
27	21	DC RNLS#21 when personal data was processed for compliance with legal obligation in union or member state law to which controller is subject.
28	21	DC RNLS#21 when collected in relation to information society services.
29	22	DC informs other controllers about RNLS#22.
30	29	DC consider available technology and cost of implementation for RNLS#29.
31	29	DC store links, copies, replications to RNLS#29.
32	21	DC limit RNLS#21 when processing for exercising the right of freedom of expression and information.
33	21	DC limit RNLS#21 when processing by union or member state law to which controller is subject.
34	21	DC limit RNLS#21 when performing task carried out in public interest.
35	21	DC limit RNLS#21 when exercising of official authority vested in DC.
36	21	DC limit RNLS#21 when necessary for public interest - public health.
37	21	DC limit RNLS#21 for archiving purposes in public interest, scientific, historical, research purposes, statistical purposes.
38	22,37	RNLS#22 likely render impossible or seriously impair the achievement of the objectives of that processing then RNLS#37.
39	21	DC limit RNLS#21 when establishing, exercising, defending legal claims.

Table B.3.: RNLS of GDPR Art. 17.

C. Semantic Models of Referenced Articles by Data Portability GDPR Art. 20

This appendix summarizes the semantic models of articles which are referenced by GDPR Art. 20 data portability.
Table C.1 shows the semantic models for GDPR article 6 paragraph 1 a and b.

Semantic Models Art. 6 (1) a,b

#	activity	actor	action	object	target / purpose
15	right: DC	DC	process	personal data	
16	right: DS	DS	provide	consent	DC / for processing for one or more purposes
17	right: DC	DC	process	personal data	fulfill contract with DS
18	right: DC	DC	process	personal data	enable DS to enter contract

Table C.1.: Semantic models of GDPR Art. 6 (1) a, b.

Another article referenced by data portability in the GDPR is Art. 9 paragraph 1 and 2 a. The semantic models for this article are summarized in table C.2.

Semantic Models Art. 6 (1) a,b

#	activity	actor	action	object	target / purpose
19	right: DC	DC	process	personal data	racial, ethic origin, political opinions, religious, philosophical beliefs, trade union membership, genetic, biometric data, health, sexual orientation
20	right: DC	DC	requires	explicit consent	to process special categories of personal data (activity#19)
21	right: DS	DS	gives	explicit consent	DC / to process special categories of personal data (activity#19)

Table C.2.: Semantic models of GDPR Art. 9 (1) (2) a.

Finally, the semantic models of article 17 are shown in table C.3.

Semantic Models Art. 6 (1) a,b					
#	**activity**	**actor**	**action**	**object**	**target / purpose**
22	obligation: DC	DC	erase	personal data	concerning DS
23	right: DS	DS	request	data deletion	DC
24	obligation: DC	DC	erase	personal data	if no longer necessary in relation to purpose it was stored
25	right: DS	DS	withdraws	consent	DC / processing of personal data
26	right: DS	DS	objects	processing pursuant	erasure of personal data
27	obligation: DC	DC	erase	personal data	unlawfully processed
28	right: DC	DC	erase	personal data	in compliance with legal obligation in union or member state law where DC is subject to
29	obligation: DC	DC	erase	personal data	collected in relation to information society services
30	obligation: DC	DC	informs	personal data for erasure	other controllers / consider available technology, implementation cost
31	obligation: DC	DC	store	links, copies, replications	other controllers / handling personal data for erasure
32	right: DC	DC	prohibit erasure	personal data	DS / processing for exercise of right of freedom of expression and information
33	right: DC	DC	prohibit erasure	personal data	DS / processing by union or member state law to which controller is subject
34	right: DC	DC	prohibit erasure	personal data	DS / performing task carried out in public interest
35	right: DC	DC	prohibit erasure	personal data	DS / exercising official authority vested in controller
36	right: DC	DC	prohibit erasure	personal data	DS / public interest - public health
37	right: DC	DC	prohibit erasure	personal data	DS / archiving purposes in public interest
38	right: DC	DC	prohibit erasure	personal data	DS / establish, exercise, defend legal claims

Table C.3.: Semantic models of GDPR Art. 17.

Bibliography

[1] T. D. Breaux and A. I. Antón, "Analyzing Goal Semantics for Rights, Permissions and Obligations," in *13th IEEE International Conference on Requirements Engineering (RE'05)*. IEEE, Aug. doi: 10.1109/RE.2005.12 [Accessed: 2018-07-17]

[2] ——, "Deriving semantic models from privacy policies," in *Proceedings of the Sixth IEEE International Workshop on Policies for Distributed Systems and Networks*, June 2005. doi: 10.1109/POLICY.2005.12 [Accessed: 2018-07-17]

[3] R. Schütte, *Die neuen Grundsätze ordnungsmäßiger Modellierung*. Wiesbaden: Gabler Verlag, 1998. [Online]. Available: https://link.springer.com/book/10.1007/978-3-663-10233-5 [Accessed: 2018-07-08]

[4] D. L. Moody, G. Sindre, T. Brasethvik, and A. Solvberg, "Evaluating the quality of information models: empirical testing of a conceptual model quality framework," in *25th International Conference on Software Engineering*, May 2003. doi: 10.1109/ICSE.2003.1201209 [Accessed: 2018-07-10]

[5] J. Riley, *Understanding Metadata: What is Metadata and what is it for?*, 2017. ISBN 978-1-937522-72-8. [Online]. Available: https://groups.niso.org/apps/group_public/download.php/17443/understanding-metadata [Accessed: 2018-06-17]

[6] B. Matthews, "Semantic Web Technologies," *E-learning*, vol. 6, Jan. 2005. [Online]. Available: https://www.researchgate.net/publication/30408878 [Accessed: 2018-06-20]

[7] The Open Group, *ArchiMate® 2.1 Specification*, 2013. [Online]. Available: http://pubs.opengroup.org/architecture/archimate2-doc/chap03.html [Accessed: 2018-06-28]

[8] S. Buckl, F. Matthes, S. Roth, C. Schulz, and C. M. Schweda, "A Method for Constructing Enterprise-wide Access Views on Business Objects," *Informatik 2010: IT-Governance in verteilten Systemen (GVS 2010)*, pp. 279–284, Jan. 2010.

[9] J. P. Almeida, M. Van Sinderen, L. F. Pires, and D. Quartel, "A systematic approach to platform-independent design based on the service concept," *Proceedings - 7th IEEE International Enterprise Distributed Object Computing Conference*, 2003. doi: 10.1109/EDOC.2003.1233842 [Accessed: 2018-09-09]

[10] M. Neppelenbroek, M. Lossek, and R. Janssen, "Twitter: An Architectural Review," *Scholars paper on Software Architecture at Utrecht University*, 2011. [Online]. Available: http://www.timdeboer.eu/paper_publishing/Twitter_An_Architectural_Review.pdf [Accessed: 2018-10-14]

[11] M. Kischporski, "EDI - Digitalisierung und IT-Wertbeitrag konkret umgesetzt," in *essentials*. Wiesbaden: Springer Fachmedien, 2017. doi: 10.1007/978-3-658-19051-4 [Accessed: 2018-09-09]

[12] Data Transfer Project, "Data Transfer Project Overview and Fundamentals," July 2018. [Online]. Available: https://datatransferproject.dev/dtp-overview.pdf [Accessed: 2018-09-11]

[13] E. Mansour, A. V. Sambra, S. Hawke, M. Zereba, S. Capadisli, A. Ghanem, A. Aboulnaga, and T. Berners-Lee, "A Demonstration of the Solid Platform for Social Web Applications," in *WWW' 16 Companion: Proceedings of the 25th International Conference Companion on World Wide Web*, 2016. doi: 10.1145/2872518.2890529 [Accessed: 2018-10-22]

[14] R. Brombacher, A. Hars, and A.-W. Scheer, "Informationsmodellierung," in *Handbuch Informationsmanagement*. Wiesbaden: Gabler Verlag, 1993, pp. 143–188. [Online]. Available: http://link.springer.com/10.1007/978-3-322-82845-3_7 [Accessed: 2018-07-03]

[15] S. L. Weibel, J. A. Kunze, C. Lagoze, and M. Wolf, "Dublin Core Metadata for Resource Discovery," Sept. 1998. [Online]. Available: https://tools.ietf.org/html/rfc2413 [Accessed: 2018-06-20]

[16] "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)," *Official Journal of the European Union*, vol. L119, May 2016. [Online]. Available: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ:L:2016:119:TOC [Accessed: 2018-11-08]

[17] H. Gimpel, D. Kleindienst, N. Nüske, D. Rau, and F. Schmied, "The upside of data privacy – delighting customers by implementing data privacy measures," *Electronic Markets*, May 2018. doi: 10.1007/s12525-018-0296-3 [Accessed: 2018-10-24]

[18] A. Ahmad and P. Mykytyn, "Perceived Privacy Breach - the Construct, the Scale, and its Antecedents," in *AMCIS 2012 Proceedings*, 2012. [Online]. Available: http://aisel.aisnet.org/amcis2012/proceedings/ISSecurity/21 [Accessed: 2018-10-24]

[19] B. Berendt, O. Günther, and S. Spiekermann, "Privacy in E-commerce: Stated preferences vs. actual behavior," *Communications of the ACM*, vol. 48, April 2005. doi: 10.1145/1053291.1053295 [Accessed: 2018-10-24]

[20] A. Kolah and B. Foss, "Unlocking the power of data under the new EU General Data Protection Regulation," *Journal of Direct*, vol. 16, April 2015. doi: 10.1057/dddmp.2015.20 [Accessed: 2018-10-24]

[21] T. D. Breaux, M. W. Vail, and A. I. Antón, "Towards regulatory compliance: Extracting rights and obligations to align requirements with regulations," in *14th IEEE Interna-*

tional Requirements Engineering Conference (RE'06), Sept 2006. doi: 10.1109/RE.2006.68 [Accessed: 2018-07-05]

[22] T. D. Breaux and A. I. Antón, "Analyzing Goal Semantics for Rights, Permissions and Obligations: A Presentation Outline Towards Machine-enforceable Policies Motivations," Tech. Rep., 2005. [Online]. Available: https://www.cs.cmu.edu/~breaux/presentations/tdbreaux-re05.pdf [Accessed: 2018-08-31]

[23] C. Bistolfi and L. Scudiero, "Bringing Your Data Everywhere in the Internet of (Every) Thing: A Legal Reading of the New Right to Portability," in *30th International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, March 2016. doi: 10.1109/WAINA.2016.106 [Accessed: 2018-07-05]

[24] Council of the European Union, "Analysis of the final compromise text with a view to agreement of the General Data Protection Regulation," Dec. 2015. [Online]. Available: http://www.statewatch.org/news/2015/dec/eu-council-dp-reg-draft-final-compromise-15039-15.pdf [Accessed: 2018-09-05]

[25] J. Kühling, B. Buchner, and M. Bäcker, *Datenschutz-Grundverordnung/BDSG : Kommentar*, 2018. ISBN 9783406719325

[26] European Commission, "Impact Assessment," *Commission Staff Working Paper*, no. SEC(2012) 72 final, pp. Annex 5, 106, Jan. 2012. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52012SC0072&from=EN [Accessed: 2018-09-11]

[27] A. D. Vanberg and M. B. Ünver, "The right to data portability in the GDPR and EU competition law: odd couple or dynamic duo?" *European Journal of Law and Technology*, vol. 8, no. 1, March 2017. [Online]. Available: http://ejlt.org/article/view/546 [Accessed: 2018-05-27]

[28] I. Graef, J. Verschakelen, and P. Valcke, "Putting the Right to Data Portability into a Competition Law Perspective," *Law: The Journal of the Higher School of Economics*, vol. Annual Review, pp. 53 – 63, 2013. [Online]. Available: https://ssrn.com/abstract=2416537 [Accessed: 2018-09-10]

[29] Lewinski, S. Brink, and A. Wolff, "BeckOK DatenschutzR - DS-GVO Art. 20," *beck-online*, 2018.

[30] Article 29 Data Protection Working Party, "Guidelines on the right to data portability," no. WP 242 rev.01, April 2017. [Online]. Available: http://ec.europa.eu/newsroom/document.cfm?doc_id=44099 [Accessed: 2018-09-11]

[31] J. Bowman, "New UK Minister's Data Protection To-Do List," Oct. 2015. [Online]. Available: https://iapp.org/news/a/new-uk-ministers-data-protection-to-do-list/ [Accessed: 2018-09-10]

[32] Article 29 Data Protection Working Party, "Statement on the 2016 action plan for the implementation of the General Data Protection Regulation (GDPR)," Feb. 2016. [Online]. Available: https://ec.europa.eu/justice/article-29/documentation/opinion-recommendation/files/2016/wp236_en.pdf [Accessed: 2018-09-10]

[33] U. Gasser, "Interoperability in the Digital Ecosystem," *The Berkman Center for Internet & Society at Harvard University*, vol. 13, July 2015. [Online]. Available: http://nrs.harvard.edu/urn-3:HUL.InstRepos:28552584 [Accessed: 2018-09-10]

[34] European Union, "Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union," *Official Journal of the European Union*, vol. 55, no. C 326, Oct. 2012. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C:2012:326:FULL&from=EN [Accessed: 2018-09-11]

[35] U.S. Congress, "Health Insurance Portability and Accountability Act of 1996," Aug. 1996. [Online]. Available: https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf [Accessed: 2018-09-11]

[36] K. Marchesini and T. Noonan, "HIPAA & Health Information Portability: A Foundation for Interoperability," *Health IT Buzz*, Aug. 2018. [Online]. Available: https://www.healthit.gov/buzz-blog/privacy-and-security-of-ehrs/hipaa-health-information-portability-a-foundation-for-interoperability [Accessed: 2018-09-11]

[37] Centers for Medicare & Medicaid Services, "Adopted Standards and Operating Rules," 2017. [Online]. Available: https://www.cms.gov/Regulations-and-Guidance/Administrative-Simplification/HIPAA-ACA/AdoptedStandardsandOperatingRules.html [Accessed: 2018-09-12]

[38] Council of the European Union, "Directive 2014/92/EU of the European Parliament and of the Council of 23 July 2014 on the comparability of fees related to payment accounts, payment account switching and access to payment accounts with basic features Text with EEA relevance," *Official Journal of the European Union*, vol. 257, July 2014.

[39] A. H. Huneke and F. Gondert, "Payment Accounts Act: New rights for consumers - basic payment account, account switching help and fee transparency," *BaFIn Journal*, vol. 6, pp. 23 – 25, 2016. [Online]. Available: https://www.bafin.de/dok/8036268 [Accessed: 2018-09-10]

[40] A. H. Huneke, "Kontenwechselhilfe: Unterstützung für Verbraucher wird zur Pflicht," *BaFIn Journal*, vol. 9, pp. 20 – 22, July 2016. [Online]. Available: https://www.bafin.de/dok/8253282 [Accessed: 2018-09-10]

[41] "Telekommunikationsgesetz (TKG) § 46 Anbieterwechsel und Umzug," 2012. [Online]. Available: http://www.gesetze-im-internet.de/tkg_2004/__46.html [Accessed: 2018-09-12]

[42] Bundesnetzagentur, "Rufnummer mitnehmen." [Online]. Available: https://www.bundesnetzagentur.de/DE/Sachgebiete/Telekommunikation/Verbraucher/Anbieterwechsel/Rufnummermitnehmen/rufnummermitnehmen-node.html [Accessed: 2018-09-12]

[43] International Electrotechnical Commission, "IEV number 351-42-26: model," Nov 2013. [Online]. Available: http://www.electropedia.org/iev/iev.nsf/display?openform&ievref=351-42-26 [Accessed: 2018-07-08]

[44] H. Stachowiak, *Allgemeine Modelltheorie*. Vienna: Springer-Verlag, 1973. ISBN 3-211-81106-0. [Online]. Available: https://archive.org/details/Stachowiak1973AllgemeineModelltheorie [Accessed: 2018-07-08]

[45] U. Kastens and H. Kleine Büning, *Modellierung Grundlagen und formale Methoden*. Hanser, Carl, 2014. ISBN 9783446442498

[46] F. Teuteberg, F. Ahlemann, and G. Brune, *Ontologie-basierte Attributierung von Informationsmodellen: Grundlagen und Anwendungsgebiete*, ser. ISPRI-Arbeitsbericht. ISPRI, 2006. ISBN 9783936475265

[47] O. Lindland, G. Sindre, and A. Solvberg, "Understanding quality in conceptual modeling," *IEEE Software*, vol. 11, no. 2, pp. 42–49, March 1994. doi: 10.1109/52.268955 [Accessed: 2018-07-10]

[48] C. W. Morris, "Foundations of the Theory of Signs," in *International Encyclopedia of Unified Science*, vol. 1, no. 2. Chicago: The University of Chicago Press, 1938. [Online]. Available: http://pubman.mpdl.mpg.de/pubman/item/escidoc:2364493/component/escidoc:2364492/Morris_1938_Foundationsoftheoryofsign.pdf [Accessed: 2018-07-10]

[49] Gartner, "Metadata - Gartner IT Glossary." [Online]. Available: https://www.gartner.com/it-glossary/metadata [Accessed: 2018-07-07]

[50] M. K. Bergman, "Advantages and Myths of RDF," April 2009. [Online]. Available: http://www.mkbergman.com/483/advantages-and-myths-of-rdf/ [Accessed: 2018-06-25]

[51] R. L. Costello and D. B. Jacobs, "XML Design - A Gentle Transition from XML to RDF." The MITRE Corporation, 2003. [Online]. Available: http://sce2.umkc.edu/csee/leeyu/class/CS560/Lecture/rdf.pdf [Accessed: 2018-11-08]

[52] A. C. Kanmani, T. Chockalingam, and N. Guruprasad, "An Exploratory Study of RDF: A Data Model for Cloud Computing," in *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications*, vol. 516. Singapore: Springer, March 2017, pp. 641–649. [Online]. Available: https://doi.org/10.1007/978-981-10-3156-4_68 [Accessed: 2018-11-06]

[53] K. Coyle, "Metadata: Data With a Purpose," 2004. [Online]. Available: http://www.kcoyle.net/meta_purpose.html [Accessed: 2018-06-20]

[54] The Open Group, *The Open Group Standard: The TOGAF® Standard, Version 9.2*, 2018, vol. C182. ISBN 1-947754-11-9

[55] S. Buckl, A. M. Ernst, J. Lankes, and F. Matthes, "Enterprise Architecture Management Pattern Catalog V1.0," Tech. Rep., Feb. 2008.

[56] A. Hess, B. Humm, and M. Voß, "Regeln für serviceorientierte Architekturen hoher Qualität," *Informatik Spektrum*, vol. 29, pp. 395–411, Dec. 2006. [Online]. Available: https://doi.org/10.1007/s00287-006-0099-3 [Accessed: 2018-06-28]

[57] T. Karle, *Kollaborative Softwareentwicklung auf Basis serviceorientierter Architekturen*. Karlsruher Institut für Technologie, 2014. ISBN 9783866448728. [Online]. Available: https://books.google.de/books?id=ofchXKqo_cMC [Accessed: 2018-11-06]

[58] S. Buckl, A. M. Ernst, J. Lankes, K. Schneider, and C. M. Schweda, "A Pattern based Approach for constructing Enterprise Architecture Management Information Models," in *Wirtschaftsinformatik Proceedings*, 2007. [Online]. Available: https://aisel.aisnet.org/wi2007/65 [Accessed: 2018-10-09]

[59] E. Cariou, A. Beugnard, and J. Jezequel, "An architecture and a process for implementing distributed collaborations," in *Proceedings. Sixth International Enterprise Distributed Object Computing*, Lausanne, Switzerland, 2002. doi: 10.1109/EDOC.2002.1137703 [Accessed: 2018-09-19]

[60] S. Puchert, "EDI," *Rechtssicherheit im Internet*, pp. 153–180, 2001. doi: 10.1007/978-3-642-59548-6 [Accessed: 2018-09-09]

[61] C. Tanner and S.-L. Richter, "Digitalizing b2b business processes - the learnings from e-invoicing," in *Business Information Systems and Technology 4.0: New Trends in the Age of Digital Change*. Cham: Springer International Publishing, 2018. [Online]. Available: https://doi.org/10.1007/978-3-319-74322-6_7 [Accessed: 2018-11-05]

[62] Data Transfer Project, "Data Transfer Project." [Online]. Available: https://datatransferproject.dev/ [Accessed: 2018-09-11]

[63] W. Malcolm, "Our preparations for Europe's new data protection law," *Google Company News Outreach & Initiatives*, May 2018. [Online]. Available: https://www.blog.google/outreach-initiatives/public-policy/our-preparations-europes-new-data-protection-law/ [Accessed: 2018-09-11]

[64] N. Horn and A. Riechert, *Practical Implementation of the Right to Data Portability: Legal, Technical and Consumer-Related Implications*, Jan. 2018. [Online]. Available: https://stiftungdatenschutz.org/fileadmin/Redaktion/Datenportabilitaet/stiftungdatenschutz_abschlussbericht_Hyperlinks_20180124_01_web.pdf [Accessed: 2018-09-10]

[65] R. Brandom, "Google, Facebook, Microsoft, and Twitter partner for ambitious new data project," *The Verge*, July 2018. [Online]. Available: https://www.theverge.com/2018/7/20/17589246/data-transfer-project-google-facebook-microsoft-twitter [Accessed: 2018-09-11]