A data sharing system architecture for scientific purposes for a Dutch health care environment

By: Chaïm van Toledo
Student number: 4237471

Version: 20171212

Supervisors: Vincent Menger and Marco Spruit

Index

List of figures

List of tables

List of acronyms

API	Application Programming Interface
CBS	Centraal Bureau voor de Statistiek (Statistics Netherlands)
CIM	Clinical Information Model
CSV	Comma Separated Values
DSM	Diagnostic and Statistical Manual of Mental Disorders
DST	Data Science Team
EHR	Electronic Health Records
ETL	Extract Transform Load
EU	European Union
FA	Functional Architecture
FTP	File Transfer Protocol
GDPR	General Data Protection Regulation
HCP	Health Care Provider
OBO	Open Biomedical Ontologies
OWL	Web Ontology Language
PDD	Process Deliverable Diagram
PEP	Polymorphic Encryption and Pseudonymisation
PSI	Private data Sharing Interface
RDF	Resource Description Framework
SSA	Software System Architecture
TLS	Transport Layer Security
UMCU	Universitair Medisch Centrum Utrecht (University Medical Centre Utrecht)
UML	Unified Modelling Language
UMLS	Unified Medical Language System
VPN	Virtual Private Network
Wbsn-z	Wet gebruik burgerservicenummer in de zorg (Use of citizenship service number in health care act)
WGBo	Wet Geneeskundige Behandelingsovereenkomst (Medical Contract Act)
Wkkgz	Wet kwaliteit, klachten en geschillen zorg (Quality complaints and disputes in health care act)

Abstract

Data sharing between researchers in different health care organisations can contribute to more insights and better scientific results. Nevertheless, sharing data between organisations is not as simple as it sounds. The patient is the data owner and regulations, both European and Dutch, restrict organisations, since they can't just give data away to another organisation.

This thesis identifies two strategies for data sharing, namely with consent of the patient or after anonymization of the data. To share detailed data about patients with a third party, the patient must be fully informed and must give her/his consent. To achieve this, the patient must understand its data. For the understanding of data, metadata can help. With already in use ontology libraries, the patient can read and understand its own data. To anonymise the data, this thesis argues, the best solution is to contract this transfer and bind the third party to what it can and cannot do with the data. In that sense, the data set can be more loosely anonymised, which contributes to a better utility for the researcher.

The main contribution of this thesis is that it gives a modelled architecture system to let health care organisations share their patients' data with other research organisations. The system takes data transfer, semantics and law into account. This thesis provides health care organisations a route map to construct a data sharing system, which is validated by experts. The validation of this thesis happened by interactive sessions, where experts could bring input and discuss the various aspects of the proposed architecture. Together with the experts, we optimised the system.

1. Introduction

Many health care organisations collect their own data and store this data in their own repositories. In these repositories, a lot of information and hence knowledge is hidden and sometimes information is discovered, a lot of times not. Nowadays health care organisations want to analyse this so called Big Data, for research purposes and for health care improvements (Ten Kate, 2016). Imagine if those organisations work together and connect their data with each other: the sum of the information outcome would be greater than when they are standalone. With data sharing, different repositories can be connected to other repositories. Researchers can get more insights between various health care providers.

For health care organisations collaborating with each other to share their patient's data, a systematically architecture is needed to fit processes into the current data system of the organisation. This architecture must apply to law and regulations and it must work with existing systems and it must be understandable. Different viewpoints and perspectives are needed to address concerns of the various stakeholders.

Laws come and go and that can sometimes change the way of working. Such an occurrence has been revealed, namely with the introduction of the General Data Protection Regulation (GDPR) by the European Union (EU). It is already in force and by 25 May 2018 it will replace the national implementations of the old directive. This new regulation will change the scope on how to treat personal data and the sharing possibilities. Interesting is how this GDPR relates to current Dutch national medical treatment laws (EU GDPR Portal, 2017).

This study will contain a case study about data sharing at a health care organisation in the Netherlands. It provides the steps and an architectural foundation that we find necessary to let a patient share its own data with a third party in a safe and legal way. A new method will be explained for data sharing in the EU from a Dutch perspective. Thereby we focus on privacy of the individual. Important hereby is that we comply to the new GDPR and Dutch medical laws.

Parts of the proposed architecture are derived from the data integration field; whereby different sources are combined. In this field, semantics are a possible solution for the understanding of the data. Everyone knows the recipes from their doctor with some scrapes and pencil strokes for some medication. But reading those papers can be hard: "What is the medicine exactly?" The patient has a clue for what it is, but still, the patient

doesn't have a full understanding unless they are medical schooled. With semantics and their practical ontology formats, patients can maybe have a clue and understand their data which gives a description about them self.

With the arrival of the GDPR in mind, it is important to administer the privacy threats of personal data sharing. The whitepaper of Verheul et al. (2016) shows a solution to share data with external parties. The solution is the PEP framework (polymorphic encryption and pseudonymisation), which can save data encrypted from physicians or devices. With the permission from the patient, data can be shared with third parties. A polymorphic key system can derive data from the framework for multiple parties in different formats. However, the product is not yet available. PEP looks also too much as a standalone solution with the focus on encryption. The product seems to be hard to integrate into the current data processing environment of a big health care organisation. The reason for this, is that the framework has its own database. Health care organisations, in most of the times, already work with a database system. The focus on data encryption is very good, but this paper argues that the whole process is much more than that. Patients must give an informed consent and understand what their data means.

With the previous findings in mind, our research question is: How, in a psychiatric health care organisation, can data sources with privacy concerns be shared with other like-wise organisations? Like-wise organisations can be other health care organisations, research institutes and other non-commercial organisations. With this thesis, hopefully, health care organisations can better share patient data for scientific research. A system will be modelled how to establish data sharing along other scientific organisations. Based on some prejudices and background knowledge, this question is divided into sub-questions, which must be answered:

- Which privacy regulations apply for this case study?
- How to give consent?
- How can we protect the privacy of the clients?
- How can we transfer data from one place to another?
- How can patients understand their data?
- How can clients choose what to share with whom?
- How can we implement the new data sharing workflow into the current?

Of the previous sub questions, five will be answered in the literature section, two of them are answered in the results, but will together with the main question summarised in the conclusion. The next part, the methods, will elaborate on how this study will be conducted. This thesis uses two methods, the first, the design science method for how this study is done. The other method, software system architecture, for how to deliver and construct the artefacts. The third part is literature and provides a brief state of the art for different aspects of the proposed system. The main architecture is modelled in section 4, data sharing system architecture and give different viewpoints with models and descriptions of how this system can be raised. In section 5, the validation, main contributions and limitations are discussed. In the conclusion section, the main question with sub questions will be answered.

The research is done at the University Medical Centre Utrecht (UMCU), situated in Utrecht, the Netherlands. The UMCU provides jobs for almost twelve thousand people and has a revenue of 1.1 billion euro's. The UMCU houses six health care related centres, including the brain centre. This centre has five specialties and this study is done at one of them, namely psychiatry. At psychiatry, a data science team tries to answer questions from the workspace and provides new insights in the bulk of the collected data (UMC Utrecht, 2017).

2. Methods

The methods section describes how this research is conducted. This paper uses a design science methodology and underneath a software system architectural (SSA) approach. The first section will elaborate the design science method. The second section will elaborate on the SSA approach. Both methods have similarities, the SSA approach section shows how it will fit in the design science methodology. The third part explains the evaluation methods. The last part reveals how the literature study is done.

2.1. Design science methodology

Hevner (2007) describes the design science approach as three cycles in three bases. There is (1) an environment base, (2) a design science research base and (3) a knowledge base. Those bases are connected through three cycles. There is (1) a relevance cycle, (2) design cycle and (3) a rigor cycle. The environment is represented by the UMCU, patients of the UMCU and third parties like the CBS or Trimbos institute. At the design science research base lies the

upcoming artefact and the evaluation about the artefact. The artefact is the modelled architecture to make data sharing possible.

The relevance cycle is between the environment and the DS research base and will evaluate the constructed artefact with the environment. This architecture will thus be evaluated by experts and by the stakeholders. Two types of qualitative evaluation methods are used: a walkthrough session with stakeholders and a presentation review for the experts. The relevance cycle is elaborated in section 2.3 Evaluation.

This thesis will use the literature review to engineer the architecture for privacy preserved data sharing in a health care organisation. The enforcement of the method is, due to practical and time limitations, out of the scope of the project. The result is therefore a proof of concept.

2.2. Software System Architecture: stakeholders, viewpoints and perspectives

A SSA approach can help to construct software. SSA is bigger than only creating software, the approach considers processes, data flows, security and so forth. Three concepts are important in the software architecture approach: stakeholders, viewpoints and perspectives (Rozanski & Woods, 2011). The stakeholders are the drivers for the project and therefore, the project needs to consider their different concerns (or requirements). This project takes three major stakeholders into account: The Health Care Provider (HCP), the third party and the patient. The HCP facilitates the data. The patient hosts its data at the HCP. The third party wants to do research with the patient's data which is stored at the HCP.

The other concept is viewpoints. "A viewpoint is a collection of patterns, templates, and conventions for constructing one type of view. It defines the stakeholders whose concerns are reflected in the viewpoint and the guidelines, principles, and template models for constructing its views" (Rozanski & Woods, 2011, p. 36). A viewpoint consists one to many views. "A view is a representation of one or more structural aspects of an architecture that illustrates how the architecture addresses one or more concerns held by one or more of its stakeholders" (Rozanski & Woods, 2011, p. 34). Seven viewpoints are given by Rozanski and Woods: (1) Context, (2) functional, (3) information, (4) concurrency, (5) development, (6) deployment and (7) operational.

A perspective is different than a viewpoint and can affect multiple viewpoints. The perspectives are more related to quality properties than stakeholder requirements. An

example hereby is security, which is a quality rather than a requirement. According to Rozanski and Woods a perspective is “a collection of architectural activities, tactics and guidelines that are used to ensure that the system exhibits a particular set of related quality properties that require consideration across a number of the system architectural views” (2011, p. 47).

Rozanski and Woods describe a whole catalogue on viewpoints and perspectives, but also state that not all of them are needed and that only what is needed, is enough. But how to determine what is needed? The SSA points out that the architectural definition need to come from the concerns of the stakeholders. In total, the architecture definition activities are in six to eight activities done: (1) Consolidate inputs; (2) identify scenarios; (3) identify relevant architectural styles; (4) produce candidate architecture; (5) explore architectural options and (6) evaluate architecture with stakeholders. If the outcome of the architecture evaluation with the stakeholders (activity six) comes with many changes, seven and eight comes into place: (7) rework architecture and (8) revisit requirements (Rozanski & Woods, 2011, p. 93).

The first three are described in this method section, namely at sections 2.2.1, 2.2.2, 2.2.3. The other three, produce a candidate architecture and the exploration of architectural options are elaborated in the results section in form of the different views. The outcome of the evaluation is in the last part presented in the results section. Activity seven and eight will also be processed in the results part. In this sense, the SSA fits good into the design science methodology, where interaction with the environment is necessary to construct a solution for a problem.

The next three sub sections will elaborate how the SSA approach will be executed and how the architecture will be constructed, these parts are: Consolidate inputs; identify scenarios and identify relevant architectural styles. The latter steps of the SSA (produce candidate architecture; explore architectural options and evaluate architecture with stakeholders) will be executed in section 4, data sharing system architecture.

2.2.1. Consolidate inputs

As first step of the design science and SSA approach, this section deals with the consolidating inputs, problem identification and therefore the motivation.

This research takes place at the data science unit of the psychiatric department at the UMCU and thus it is a case study (n=1). One of their goals is to get more insight in saved and combined data, with data created at the UMCU, but also from other organisations, for example the Dutch governmental statistical office Centraal Bureau voor de Statistiek (CBS) or other psychiatric organisations (Jongejan, 2016). The data science team (DST), under the supervision of the head psychiatrist, helps the psychiatric department to analyse health care problems, answer questions from the workspace, tries to improve treatments and give new insights to the employees. The DST also wants to combine multiple data sources from other public funded Dutch organisations. For example, to analyse patients records with the micro data services of CBS. Juridical problems arise when data proliferate out of the health care environment, such thing happens with the previous explained case of the CBS.

The DST works in a data mart, where data is pseudonymised from the electronic health records (EHR). No database is yet used: data is currently stored in comma separated values (CSV), SPSS and other kind of formats. The data mart situation is sketched in figure 1. The figure shows the basic outline of the current system. With the extract, transform and load (ETL) processes, data is derived from HiX (the EHR system), which is located at the internal sources. From the ETL processes, the data is staged (see Staging box) on a secured and restricted server. From there, researchers analyse the data for their own projects or for other employees. At last, the DST create an output, like interactive diagrams, reports or presentations. From there, the researchers get possible new questions from the working floor to answer. The dotted lines in figure 1 are future aspirations and shows the potential collaboration with the CBS and to load more data from other organisations in their ETL processes. The other dashed arrow shows the share of data with external parties.

But in short, what do the stakeholders want (HCP, third party and patient)? The HCP wants to share their collected data with other parties to improve health care. They also want to do it legally and safe. The patient wants control over their data and know what happens with their data. The third party wants to enrich their research data with the use of external data. In this case study, the UMCU is the HCP. The patient receives their treatment at the HCP. The focus in this paper is most on the HCP, for them an architecture should be to be implemented in the current data process. The method must obey the newest privacy laws. In Europe, the GDPR is in effect, but there are also national laws which (can) influence the method.

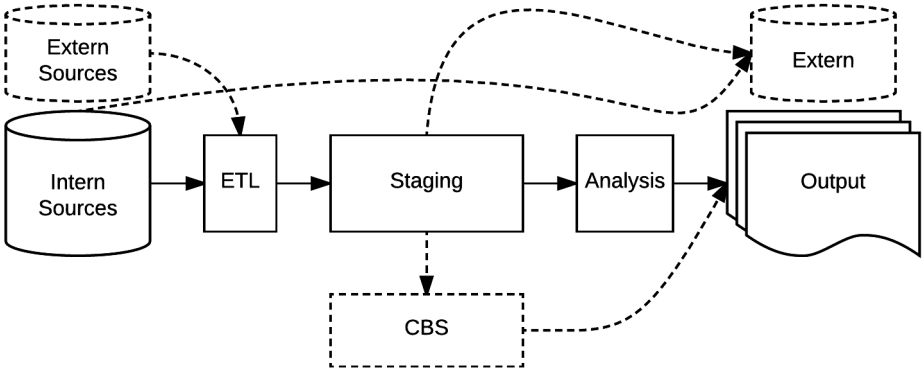

Figure 1: Current situation with future aspirations: the dotted lines are the aspirations

2.2.2. Identify scenarios

In this part, scenarios and relevant architectural styles are defined. As mentioned in 2.2.1, this thesis takes three stakeholders into account: (1) HCP, (2) patient and (3) third party. For identifying scenarios, the functional scenario format is used (Rozanski & Woods, 2011, p. 150): Overview (brief description), system state (before scenario), system environment, external stimulus (cause of scenario) and required system response. Three main scenarios are derived and written in table 1. These scenario's must lead to the needed viewpoints and perspectives.

Overview	**System state**	**System environment**	**External stimulus**	**Required system response**
Request for data sharing.	Nothing	E-mail box	Third party wants to get some sort of data and send a request.	Admins needs to check third party, purpose, what kind of data. Then give a positive or negative sign.
Request for consent.	System give signal for new request. In the request the purpose of the research and	Patient portal	Approved third party data request.	Patient response with its consent.

	which data will be shared.			
Pull data.	Data is put down in a data place, where the data can be used for research.	Servers with connections to outside the environment.	Data needs to get downloaded for research.	Data is downloaded.

Table 1: Functional scenarios

2.2.3. Identify relevant architectural styles

Three main scenarios are described in table 1. This will lead to a short analysis for using viewpoints, described in table 2. The viewpoint gives different insights in the architecture. Table 3 describes which perspectives are necessary to add to the different viewpoints. In short, the relevant viewpoints are: context, functional, information and operational. The perspectives are: usability, internationalisation, regulation and security.

Viewpoint	**Relevancy**
Context	A context viewpoint is necessary to describe relationships and interactions between system and stakeholders. This viewpoint is helpful to check the scope of the architecture and responsibilities. The context viewpoint is helpful to get an understanding of most of the system's stakeholders. This viewpoint will be used because of the fast understanding of the system.
Functional	"Functional viewpoint describes the system's runtime functional elements, responsibilities, interfaces and primary interactions" (Rozanski & Woods, 2011, p. 244). This viewpoint shows how the different functions work with each other. It is a viewpoint that is helpful to construct the system and will be used in this architecture.
Information	The information viewpoint is a necessary viewpoint for the reason of how the data will flow through the system and complements the functional viewpoint of how the data will likely be flowing through the system.
Concurrency	The concurrency viewpoint is not necessary. Concurrency can be

	helpful to coordinate concurrent processes, but to reflect to the previous table, there are not that many stakeholders doing something in the system at the same time. The concurrency viewpoint will also be handled by the information viewpoint with the use of a process-deliverable diagram. Therefore, it will not be handled in this architecture.
Development	The development viewpoint supports the software development process. How the system is constructed, is out of the scope of this research. Technical obstacles can eventually be dealt with within perspectives or suggestions at the operational, deployment or functional viewpoint. Thereby the construction of the system can happen on different ways on different operating systems.
Deployment	The deployment viewpoint will describe the system in its environment, such as interactions with hardware, the technical environment or other operating systems. In this study, the deployment is used to set the proposed system in the UMCU's technical environment.
Operational	The operational viewpoint can be helpful for the reasons of how to handle the operational data flow. And what must happen to monitor a good data sharing process. The focus can also be on how to install and manage the system, but in this case the operational viewpoint is used to see how the environment is using the system through a scenario. Then it will be analysed with an administration model.

Table 2: Relevant viewpoint analysis, viewpoints which are underlined are used in the architecture

Perspective	**Description**
Usability	The usability perspective will be used to ease how patients will operate with the system. This perspective addresses this issue.
Internationalisation	The internationalisation perspective will be combined with the usability perspective, not to overcome a language problem, but terminology understanding problems. This perspective is thus not only needed for people who cannot read or speak the language, but

	also to overcome cultural differences.
Regulation	The regulation perspective is helpful to point out risky parts in the system.
Security	The security perspective is needed, because of the working with sensitive information.

Table 3: Relevant perspective analysis

2.3. Evaluation

For testing and evaluating the proposed architecture, this study first follows the methods described by Venable, Pries-Heje and Baskerville (2012). The framework to find an evaluation method is divided in two dimensions. The first dimension is *artificial* evaluation versus *naturalistic* evaluation. Artificial evaluation is, as the name suggests, for only technical artefacts. The artificial evaluation "includes laboratory experiments, field experiments, simulations, criteria-based analysis, theoretical arguments and mathematical proofs" (Venable, Pries-Heje, & Baskerville, 2012, p. 430). Naturalistic evaluation can happen within the organisation, with real people and is for artefacts which need to be placed in the real world. Naturalistic evaluation includes "case studies, field studies, surveys, ethnography, phenomenology, hermeneutic methods and action research" (Venable, Pries-Heje, & Baskerville, 2012, p. 430). The second dimension is *ex ante* versus *post ante* evaluation. When there is an uninstantiated artefact (like a design or model), ex ante evaluation comes in place. Post ante evaluation comes in place when there is an instantiated artefact.

With these dimensions in line, this paper uses a combination between artificial and naturalistic evaluation and an ex ante evaluation. In this combination action research, case study, focus group, participant observation, survey and simulations like a walkthrough are represented. Out of these evaluation methods, a cognitive walkthrough session and expert reviews are used.

A guide to cognitive walkthrough session is provided by Polson, Lewis, Riaman and Wharton (1992). The walkthrough will simulate how the system can be used. In this way, the user interaction with the system will be tested. The evaluation method will check if the flow through the system is correct.

Another evaluation is performed by presenting the artefact to experts. The experts will review the artefact through a presentation and make recommendations to get to the

correct artefact. The advantages of a presentation are that a presentation can be quickly executed and feedback is given immediately from the audience (Rozanski & Woods, 2011). This paper chooses a variety of experts, to review the architecture from different 'expert' angles.

2.4. Literature study

The foundation of the modelled architecture is the literature review. The literature provides solutions and insights and will help to shape the proposed solution (Kitchenham & Charters, 2007). The literature review itself, is executed by a systematic literature review (SLR). Reasons for a SLR are to summarise the current state of the art: what is already proven and what needs to be examined. Three stages are necessary for a SLR, namely: (1) the planning, (2) conducting and (3) reporting phase.

In the planning phase, several terms are used, which are based on the scenarios of table 1 and what this paper calls environmental gathering of terms: terms that are gathered by asking the environment. The inventory of terms by asking the environment, fits in the design science methodology. With the conduction of the SLR, the articles for the regulation are scanned on if they are in the context of health care or research. The used terms are listed in table 4.

Besides the SLR, some articles and sources are 'snowballed'. This means that some of the articles were not found at Google Scholar by using the search terms, but by looking at the literature list of an article (Wohlin, 2014). In the field of the regulations, the sources of the articles (the actual laws) are snowballed, this is the case for the regulation itself.

Finally, some of the insights of the literature exploration are gained through sessions with experts or simply following newspapers, for example the case of the Virtual Private Network (VPN) remote way for connecting third party to HCP, which described in section 3.3.3. The section refers to the website of the CBS, but it originates from a meetup session about legislation and CBS Microdata Services at the Trimbos Institute. The description of the PEP-framework at section 3.3.2 is discovered by following newspapers who reported on this piece of software.

For the reporting phase, the current state is written down, however, for the privacy laws and regulations, extra support of experts is used. These experts are two information security specialists and one lawyer in the field of use of data in health care.

Keywords	Artefacts
General Data Protection Regulation (+ medical research + privacy)	Scientific papers of (Chassang, 2017), (Rumbold & Pierscionek, 2017) about the implementation of the GDPR. Documents and websites of (Autoriteit Persoonsgegevens, 2017), (EU 2016/679, 2016) about data protection in the EU and in the Netherlands (PrivacySense.net, 2015). Correspondence with the legal department of the UMCU about what can be done and what not.
Semantic interoperability	Scientific papers of (Gardner, 2005), (Dustdar, Pichler, Savenkov, & Truong, 2012), (Gomez-Cabrero, et al., 2014) (Heiler, 1995), (Fortineau, Paviot, & Lamouri, 2013), (IDABC, 2004), (Moreno-Conde, et al., 2015), (Sharda, Delen, & Turban, 2014), (Chang & Terpenny, 2009), (U.S. National Library of Medicine, 2016). These articles are used in the sense of using data management.
Ontology + data integration	Scientific papers of (Doerr, 2003), (Gardner, 2005), (Slater, Bouton, & Huang, 2008), (Hastings, Smith, Ceusters, Jensen, & Mulligan, 2012).
Transport Layer Security	Scientific paper of (Bhargavan, et al., 2014). For Transport Layer Security, the need was just for an overview of the current state and not for proving new methods in this field.
Anonymisation	Scientific papers of (Toledo & Spruit, 2016), (Gambs, Killijian, & del Prado Cortez), (Rumbold & Pierscionek, 2017)
Consent	Scientific paper of (Alderson & Goodey, 1998). Website of (PrivacySense.net, 2015).

Table 4: Used terms with founded articles, documents and websites

3. Literature

Next part will elaborate the current state of the art of the many aspects of the architecture, which will give different insights how to produce different aspects of the proposed artefact. The first part begins with the strategies of data sharing, whereby the zoom is on anonymisation and giving consent. The second part elaborates the European and Dutch legal

framework of using data and sharing it. At the third part, the focus is on how data can be connected in a technical perspective: how to get data from A to B. The fourth and last part is about semantics and is intended to gain insight into what data means and how to understand it.

3.1. Data sharing strategies

In health care research in the public interest there are two strategies of data sharing: by anonymisation of the data set or to ask for consent at the patient. In this section, the first part is about how to anonymise data and how to know if something is anonymised. The second part is about the different views of giving consent.

3.1.1. Anonymisation

In the case of anonymisation, there are three enablers for anonymisation in database environments: metrics for the measurement of anonymisation, query restriction and data perturbation options (Toledo & Spruit, 2016). The first and third are helpful in case of exclusive data sharing: with data perturbation, the database can comply to predefined anonymisation metrics. Query restriction means that the user cannot do all the queries in a database.

All kind of measurements like k-anonymity, p-sensitive, i-diversity (or better, the k-anonymity family) or differential-privacy got their critiques (Toledo & Spruit, 2016, p. 3). Individuals were traced back at the so called anonymous data sets, like for example the Netflix case (Gambs, Killijian, & del Prado Cortez; Narayanan & Shmatikov, 2006). In this Netflix case, researchers used the Netflix movie ratings of 500.000 users. Combined with the IMDB.com as background knowledge, the researchers could uncover “their apparent political preferences and other potentially sensitive information” (Narayanan & Shmatikov, 2006, p. 1).

On the other hand, the stronger the privacy measures, the further the utility will drop. If, for example the privacy measurement works with i-diversity, the data set is in total grouped with other likewise patients, the way to look for differences in populations is hard to find. Differential-privacy could be a potential solution to tackle the utility vs privacy dilemma. It adds random noise to the data set on a statistical way (Green, 2016). However, it

is difficult to implement and it looks like there are no production systems available to automatize these tasks. From the Harvard University, the PSI (Private data Sharing Interface) tool is introduced. It let researchers upload private data to the system, decide what kind of statistics they would like to use and "release privacy preserving versions of those statistics to the repository" (Dataverse, 2016). But it seems that the tool is still in prototype phase and what is also important, is that the tool is probably located in the United States. Which means that the data should travel outside the borders.

But this case is different than for example the Netflix case. Data would be shared with other researchers at other organisations and there lies the solution: with proper assurances and safety measures by both organisations with proper contracts, researchers will not make attempts to identify persons in the anonymous dataset, because they are bounded to a contract (Rumbold & Pierscionek, 2017). What is also important, is not to openly publish the data sets. In this sense, the data may be stored for a fixed time set, for what is reasonable for both the researcher as for the data set publisher.

3.1.2. Consent

Consent is a "voluntary agreement to or acquiescence in what another proposes or desires; compliance, concurrence, permission" (OED Online, 2017). McGuire, et al. (2011) distinguish three types of consent in the field of data sharing in genome research: traditional, binary or tiered. Traditional means that the person has no choice but sharing all their data to both publically and restricted research databases, otherwise it doesn't participate in the project. Binary means that the person participates in the research, but has the option to put its data in the public and restricted research databases. With the tiered option, the person can participate in the research, and have the choice where the data is stored: public, restricted or both.

Another distinction is made by Alderson and Goodey (1998): informed consent, voluntary consent, consent to research and competent consent. With an informed consent, the participant will know: the nature and the purpose of why their data will be shared, what kind of risks, harms and possible benefits will be outcome, what the alternatives are, and the intended effects and eventually side-effects. Voluntary consent means that there is no form of constraint or coercion to participate, the participant knows that refusal or withdrawal is an option that would not affect her/him, thereby the participant can negotiate and ask

questions about the participation. With consent to research, the participant knows how the research is conducted, where the researchers are hoping for, which people are in the research team and what risks and harms can happen when participating. Competent consent means that the participant can make an informed decision.

PrivacySense.net (2015) on the other hand, describes a rather action driven way of consent. They distinguish three types of consent: Explicit, implicit and opt-out consent. At explicit consent the participant has the option to agree or disagree with the clearly presented way of data sharing. Implicit consent means that a participant voluntary gives its personal data. The benefit is in most of the cases mutual. In this case, it seems to be 'logic' that the data is shared with the organisation. Opt-out consent means that consent is granted, until the participant says no.

3.1.3. Conclusion

This part shows two ways for data sharing, namely anonymisation and by giving consent. The first part, anonymisation, answers the sub question: How can we protect the privacy of the clients? The answer shows also that anonymisation can be hard. Through other openly data sets the individual can be recognised (see Netflix case). The anonymisation part shows us that adding measurements, aggregation and noise is needed to anonymise the individual and together with a good contract, these linkages and openly publications of the data can be prevented. The benefit of this method is that data will be easily shared with other parties. However, the utility of the data set drops. The second part, giving consent, is the other way to share data. That part answers the sub question: How to give consent? The subject can give an informed consent to let their data be shared with another party. There are also forms of consent, like opt-out, but it is not likely that it can be used for data sharing very personal data. The benefit of the informed consent method is that the utility increases, because of the detailed data of the individual is much likely enriched. The downside of this method is that everyone in the data set must be asked if they give consent for data sharing. This is a time-consuming process, with eventually traceability problems on the horizon. With a lot of non-consents, the data set also tends to be smaller.

3.2. Legal framework of personal data sharing

This part will elaborate the European perspective, which gives a not complete view of the law. This incomplete view is caused by the additional national laws. The second part goes deeper in the Dutch law, because this case study is situated in the Netherlands.

3.2.1. European perspective: General Data Protection Regulation

The data sharing case at the UMCU is situated in the Netherlands. The Law of Personal data protection (Dutch: Wet bescherming persoonsgegevens), contains the most important rules concerning personal data. The Dutch Personal data protection law is derived from the European directive (95/46/EG) (EU GDPR Portal, 2017). But in the EU, the national laws and thus the European directive are overthrown on 25 May 2018 by a new European law. This new law is the GDPR and is already in effect throughout the whole EU (Rumbold & Pierscionek, 2017; De Silva, Liu, & Nabarro, 2017; Chassang, 2017). The new law applies to all data controllers (those who collect and process data) and processors (only processing data).

Important in the new law is the prohibition of processing personal data, such as genetic, biometric and health data. There are exceptions, like giving explicit consent by the subject or when the processing is necessary for medical research in the public interest (Rumbold & Pierscionek, 2017).

Giving explicit consent cannot be a silent consent. To put it simply: the organisation cannot say: “We will use your data, unless you object.” The subject must know what will happen with her/his data, for how long, for what purpose: it must be a freely given clear affirmative act (EU 2016/679). When there are multiple purposes, the subject must give consent for all of them. Interesting is recital 33, when the purpose of the processing is scientific research. The recital states that it is often not possible to identify the full purpose beforehand. The data subjects can in this matter assign their consent to broader scientific areas. Of course, the controller or processor must apply the “recognised ethical standards for scientific research” (EU 2016/679, p. 6).

About the responsibility of the controller, the law states that the controller shall implement appropriate technical measures and organisational data protection policies. An organisation can show that their measures and policies are adequate by applying codes of conduct, approved certification and the implementation of data protection by design.

Article 25, Data protection by design and default, gives a few guidelines for organisations to have in mind when data is collected and processed. These guidelines are: the amount of personal data collected; extent of their processing; period of their storage; and accessibility of the data storage. These points of the regulations are not entirely demarcated. Chassang (2017) states that the organisation must take technical and organisational 'state-of-the-art' measures to protect data. These can be measures such as pseudonymisation, encryption or anonymisation of data or in short: the implementation of privacy enhanced technologies. In this light, it is notable that privacy and data protection is an ongoing process.

The regulation also is focused on organisations that "aim to process a considerable amount of personal data [..] and which could affect a large number of data subjects and which are likely to result in a high risk" (EU 2016/679, p. 119/17). For that kind of organisations, a protection impact assessment should be done. There is an exception: "The processing of personal data should not be considered to be on a large scale if the processing concerns personal data from patients or clients by an individual physician, other health care professional or lawyer. In such cases, a data protection impact assessment should not be mandatory" (EU 2016/679, p. 119/17).

Recital 54 states that the processing of data for reasons for the public interest in public health can be without a given consent. This includes:

> "all elements related to health, namely health status, including morbidity and disability, the determinants having an effect on that health status, health care needs, resources allocated to health care, the provision of, and universal access to, health care as well as health care expenditure and financing, and the causes of mortality" (EU 2016/679, p. 119/17).

This doesn't mean that the data can be processed by third parties: "processed for other purposes by third parties such as employers or insurance and banking companies" (EU 2016/679, p. 119/11). Sharing data with other organisations still looks a bit vague, as Rumbold and Pierscionek state about the public interest: "it might exclude [..] arrangements that have no evidence of benefit sharing" (2017, p. 2).

Automated individual decision-making is also prohibited in the GDPR: "The data subject shall have the right not to be subject to a decision based solely on automated processing" (EU 2016/679, p. 119/46). Three exceptions are included: when it is necessary

for the subject and the data controller to get a contract with each other; when it is authorised by the EU or the member state; and when the data subject has given consent.

3.2.2. Dutch perspective: Medical Contract Act

In the Netherlands, the Medical Contract Act (Dutch: Wet Geneeskundige Behandelingsovereenkomst, WGBo) is in effect. The act is on top of the old Law of Personal data protection and the GDPR. The GDPR permission of not asking for consent in the medical public interest for research purposes is not applicable: “a physician or caregiver may disclose identifiable patient data to researchers when the patient authorized the disclosure” (Hoytema van Konijnenburg, Teeuw, & Ploem, 2015, p. 1575). The authorisation can be given in twofold, the treating physician can formulate what the person will do with their data under the article of 457 of the WGBo, the patient can opt-out eventually (patients can ask not to use their data for further research (PrivacySense.net, 2015)).

The Quality complaints and disputes in health care act (Dutch: Wet kwaliteit, klachten en geschillen zorg, Wkkgz) applies the need for the health care organisation to improve itself in health care (article 7). Article 9 explains that it does not need any consent of the patient to improve the health care.

Another important act, is the Use citizenship service number act in health care (Dutch: Wet gebruik burgerservicenummer in de zorg, Wbsn-z). This law restricts the use of citizenship service number (in Dutch: burgerservicenummer or BSN) in health care. Even with permission of the patient, the number cannot be shared with other organisations in the field of scientific research. Thus, for connecting patients’ data to other research organisations, a BSN is not usable, patients must be identified with other characterisations, such as name, birthday, birth place, etc.

Sharing patients’ personal data with other organisations will say that the physician–patient privilege will be broken. To share legally the personal data with other organisations, an informed consent comes in place: the patient must be completely informed what will happen with his/her data and for what purposes (Autoriteit Persoonsgegevens, 2017).

3.2.3. Conclusion

This part discusses what can and what cannot be done with data use and sharing and answers the sub question: “Which privacy regulations apply for this case study?” The short

answer is the European GDPR and three Dutch national medical laws. The use of medical personal data for better health care and research is permitted if it is for the sake of the public interest, but in the case of the Netherlands, this is applicable under the Medical Contract Act. Patients still to be informed on what an extern organisation will do with their data. The psychiatry department should take state of the art safety measures to protect the subjects. Data Protection by design and default must be the norm. Sharing data with other parties can be permitted, if the department anonymises the data or when the patient gives consent. In the case of anonymisation, the subject cannot be traced from the data. Sharing personal data with other parties can be permitted, if the subject signed a consent for a specific purpose or research field of the data. The use of data for automated individual decision making is permitted, if the subject gives consent. Table 5 gives a short overview of what is possible in the sense of research in the own environment and if it is sharable with external parties.

	Personal	**Anonym**
Use of personal identification number	Never	Never
Use of names	Mostly, not needed	Never
Use of quasi identifiers (birth, postal code etc.)	Opt-out consent	Informed consent
Share with external parties	Informed consent	Yes

Table 5: Use of personal data for research purposes in Dutch health care

3.3. Data transport

This part elaborates how the third party can connect to its wanted data. The first part elaborates how this can be established when the data is downloaded from the HCP environment. Second, the PEP framework, a not yet published framework to receive and transfer data, encrypt data and can share it to multiple third parties. The third and last part shows a VPN solution to let the data stored in the HCP environment and let the third party work in a remote area. The solution is based on the CBS microdata environment.

3.3.1. Downloading

Three safe possible strategies to download the data from the HCP to the third party were found. The first one is through a secured connection to the file transfer protocol (FTP) server of the HCP. The second strategy is by a connection party extern from the HCP. The last and third one is through a physical way of transporting data from the HCP environment to the third party with a secured USB-drive.

For downloading through a secured connection, the Transport Layer Security protocol (TLS) is the most deployed secured communications protocol (Bhargavan, et al., 2014). TLS is the successor of Secure Sockets Layer protocol. With the TLS handshake protocol, a secure connection can be created over the internet. The handshake addition ensures that when the connection is intercepted by a malicious party, the transfer is encrypted. TLS is also applicable to a FTP server, and is called a FTPS connection. The client begins with sending a TLS authentication request to the server, the server response positive. The server and client completes a TLS handshake and continue the username and password authentication and FTP interaction over the secured connection (Springall, Durumeric, & Halderman, 2016).

The second solution for safe downloading from the HCP to a third party, is to use SURFFiletransfer or a likewise product (WeTransfer etc.), which gives a solution to transfer big files from one computer to another. The service operates in the Netherlands, so for Dutch file transfers there is some certainty that the data doesn't travel outside the national borders (SURF, 2017). The solution is very easy to implement, but it is harder to scale, if for every change in the data set a new transfer must be established. An API could automate this function but there is not any function yet.

Last solution is to put the data on a secured USB-drive. In this way, the third party must come to the HCP and transfer the data from the computer to the USB-drive. The method is a bit devious, because of the travel of the third party, thereby it is not useful when there is an update to the data. To scale this method with multiple third parties is also not very useful. On the other hand, the method is a very cheap solution. Besides all that, the USB drives can also be hacked if the drive falls in the wrong hands (Kim, et al., 2013).

3.3.2. PEP-framework

The new PEP framework can also help with data sharing. The framework is now developed at the Radboud University in Nijmegen and helps patients with Parkinson's disease to share their data with other (non-)commercial organisations (Verheul, Jacobs, Meijer, Hildebrandt, & Ruiter, 2016). The system looks promising, because it aligns with the newest EU data and privacy regulations and lets users control over their own data by creating different keys to decrypt their data for sharing different kind of pseudonymised data. The PEP framework is not yet released. The date of the software release is probably this fall.

The benefit of the PEP framework is that the data can be accessed in different ways. When data is send to the storage, the data is encrypted in a polymorphic way and cannot be accessed anymore at the storage provider/database. Between the storage and the data retriever (scientists, doctors etc.), there is a transcriptor. The transcriptor plays the role of trusted intermediate party and "[..] is a central converter who exclusively knows how to turn the wheel on a polymorphic lock so that keys of specific parties fit" (Verheul, et al., 2016, p. 5). The role of the transcriptor is crucial, it connects the involved stakeholders with the database, but it does not possess any key. The access manager is the portal for the patient, after an authentication, the user can decide which of her/his data can be accessed (commercial or non-commercial organisations) and in what form (raw data or pseudonymised data). The user can also see who accessed the data. Additions, modifications and access to the storage will always be logged.

The great benefit of this all, is that when the data will be encrypted, the decision of who can decrypt the data can take place later in the process. Figure 2 shows a schematic overview of this architecture.

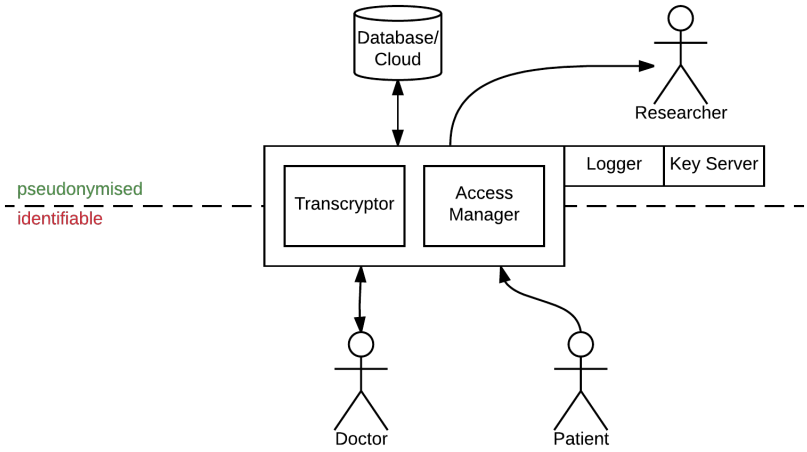

Figure 2: PEP framework architecture (inspired on the paper of Verheul, et al. 2016)

3.3.3. VPN remote environment

This part elaborates a VPN remote environment of the basis of the CBS microdata services. The CBS provides a service for scientific research organisations, whereby the researcher can extent their data set with the data of CBS under strict conditions. Via an encrypted securitised VPN, the researcher can connect to the environment of the CBS with their data and connect it to their data. The microdata of the CBS cannot be taken from the environment, only the statistical output of SPSS, R and other likewise applications can be taken.

The CBS case is interesting, because this can be arranged for the data sharing case. The data technically doesn't leave the environment; however, the physician-patient privilege still needs to be broken if someone outside the environment is working with the pseudonymised data.

3.3.4. Conclusion

The previous sections show different ways to let data safely travel from one place to another and answers the following sub question: "How can we transfer data from one place to another?" To secure the data transfer for downloading from a FTP server, the TLS handshake protocol can help, it is very scalable, but is probably harder to implement than the other two downloading methods. To transfer data with a transfer party is also a good option, especially if there is an option to automate the file transfer. A secured USB stick is also possible, this method is easy to implement, but not very scalable. The PEP-framework is a more complete

product, but unfortunately, it is not released yet. Also, it is harder to implement this framework in the current health care digital environment. The method with the VPN at the CBS, shows a different way of connecting the outside with the data. It prevents the proliferation of the data outside the HCP environment.

3.4. Defining data

The next section is about defining data. Defining data is key for understanding data. If patients do not understand the data, they cannot give an informed consent, for the simple reason that they are not fully informed. For a third party, it can also be interesting, if the data is formulated different than they expected. The first part gives an overview of the current state of semantic interoperability. Semantics can help patients to understand their data, because semantics can give a description of data. Otherwise external parties can understand what's the data about and how they can work with it more easily. The second part gives an overview about ontologies. Ontologies can function as metadata and can thus provide an easy format for the other party.

3.4.1. Semantics

Semantics are, according to the Oxford English dictionary: "the meaning of signs; the interpretation or description of such meaning; the study of the meaning of signs, and of the relationship of sign vehicles to referents" (OED Online, 2017). Semantics are an important aspect of data integration, to combine multiple data sources, to overcome semantic conflicts, so that eventually semantic interoperability can occur. Managing semantics will ensure data understanding (Gardner, 2005). Semantics have a wide area of research, for example in semantic interoperability in database research (Dustdar, Pichler, Savenkov, & Truong, 2012), but also in life sciences (Gomez-Cabrero, et al., 2014).

The field of semantic interoperability can be interesting for data sharing. Multiple data sources can be linked to each other and eventually new insights can be acquired. According to Heiler (1995), interoperability is "the ability exchange services and data with one another", "semantic interoperability ensures that these exchanges make sense, that the requester and the provider have a common understanding of the requested services and data" (Heiler, 1995, p. 271). Another definition states that "interoperability can be defined

as the ability of two systems or more to communicate, cooperate and exchange data and services, despite differences in languages, implementations and executive environments or abstract models" (Fortineau, Paviot, & Lamouri, 2013, p. 363).

Three levels can hereby be distinguished, namely the technical, organisational and the semantic levels. For an achieved good interoperability, all three levels need to be completed. Organisational interoperability is concerned with the business goals, processes and collaboration between the organisations for exchanging data. Technical interoperability concerns linking the data on a technical infrastructure level. Systems need to be enabled to share the data with each other. Semantic interoperability is about the meaning of the data and knowing what the organisations mean with their data (IDABC, 2004).

Semantics, the information structure, technological specifications and how information is organised and is described can be caught in the expression clinical information model (CIM) (Moreno-Conde, et al., 2015). In other words, CIM manages the whole information processes of storing, controlling and analysing the data. One of the missing pieces in semantic data integration of Moreno-Conde et al. (2015, p. 943) is "a unified process to guide CIM definition, including the description of best practices to increase the quality of the CIMs." However, little is told about the practical application of how to integrate the data and manage the semantics, although various papers show a non-practical semantic integrating methodology.

Another way to look at semantics, is to look at the metadata. "Metadata describes the structure of and some meaning about data, thereby contributing to their effective or ineffective use" (Sharda, Delen, & Turban, 2014, p. 69). One way to make metadata explicit, is to look at the so-called ontologies. "Ontology is a formal specification of domain knowledge and has been used to define a set of data and their structure for experts to share information in a domain of interest" (Chang & Terpenny, 2009, p. 863). Ontologies can make data understandable and when two data sources have ontologies, ontology matching can easily occur to look to the similarities and then match different attributes of the two sources.

To integrate two data sources or to understand data from other health care organisations, the Unified Medical Language System (UMLS) can be useful. "UMLS integrates and distributes key terminology, classification and coding standards, and associated resources to promote creation of more effective and interoperable biomedical information

systems and services, including electronic health records" (U.S. National Library of Medicine, 2016). The next part will elaborate further on ontologies.

3.4.2. Ontologies

Ontologies are very useful in handling semantics (Doerr, 2003). "An ontology formally defines different concepts of a domain and relationships between these concepts" (Ghawi & Cullot, 2007, p. 1). "These associations between concepts are captured in the form of assertions that relate two concepts by a given relationship" (Gardner, 2005, p. 1004). So, with the use of ontologies in data integration, similarities and differences of sources can be shown, in other words, a common vocabulary enables mutual connection of the sources. Ontology formats can be written in the Resource Description Framework (RDF), but also the Ontology Web Language (OWL). The OWL language can be used for inference, which means that reasoning can be applied with theses ontologies (Slater, Bouton, & Huang, 2008).

In a single-ontology/global-as-view or hybrid approach, an existing and open ontology framework can be useful. With these standardised ontology libraries, data sources can easily be linked to existing ontology schemes. For example, the Open Biomedical Ontologies (OBO) consortium has the aim to give a hold to the proliferation of new biomedical ontologies and manage to let multiple data sources combine (Smith, et al., 2007). But there are also more specialised ontologies, such as the DSM ontology (currently DSM-5): "DSM provides not only a classification of disorders but also guidance as to the diagnostic criteria for these disorders in the form of checklists of symptoms, with counts of how many symptoms of a various sort are required for the condition to be diagnosed" (Hastings, Smith, Ceusters, Jensen, & Mulligan, 2012, p. 1).

3.4.3. Conclusion

With these sections about semantics and ontologies, the sub question "How can patients understand their data?" is answered. The question is answered by looking to semantics and ontologies. If ontologies are well implemented, the patient can more easily understand its data. The meta data can assure a bridge between the patient knowledge and the medical language. RDF and OWL can play a role to assure the bridge on a technical way. Especially, with the use of existing meta data libraries, such as DSM or OBO. DSM-5 can help with the semantic understanding of the data in the psychiatric field. Besides of the benefits for the

patient, semantic interoperability can also be helpful between organisations. Data sources can be linked in a productive environment. This mean that organisations can easily work with each other. However, for full semantic interoperability it is important to comply the three levels: technical, organisational and semantical. For looser collaboration ontologies can be helpful, because of data understanding.

4. Data sharing system architecture

In the next phase, the proposed architecture will be shown for data sharing for research purposes. The creation of the architecture is based on the SSA method described in section 2.2. The viewpoints sections give one or more views. The perspectives address quality concerns and give solutions, for example the security perspective shows potential vulnerable parts of the system and provides solutions for that vulnerable parts.

The results section is divided in five sub sections. The first section will elaborate the context viewpoint with a context and use case diagram and shows the scope of the architecture in a schematic way. The second section shows the functional viewpoint, with the security and usability perspectives. Third section elaborates the information viewpoint and shows how the information flows and where this is saved. A regulation perspective gives insight into where attention needs to be focused for getting the system running. The fourth viewpoint is deployment and shows the system in relation with surrounding systems. The deployment viewpoint is the context viewpoint extended, where the system is no longer a black box, but operating with other parts of the UMCU. The last viewpoint, the operational viewpoint, gives insights into how the system will be running through a scenario and how it should be administered.

4.1. Context viewpoint

“The context viewpoint describes relationships, dependencies and interactions between the system and its environment” (Rozanski & Woods, 2011, p. 247). In this context viewpoint, two views are used. The context diagram shows the environment and the use case scenario shows the system functions for the stakeholders.

4.1.1. Context Diagram

The context diagram, shown in figure 3, visualises the system as a black box in its environment with its interacting stakeholders. The context diagram shows, in a basic way, where the proposed architecture is placed. The figure shows the data sharing system in its environment, namely the EHR database and an ontology / metadata system and involved stakeholders. The ontology / metadata systems are standardised classification systems for a patient diagnosis. An example is DSM-5 for mental disorders, where physician can diagnose a patient with some sort of mental illness classification.

The three main stakeholders are not specified involved with the system. The data sharing system needs the EHR database to share the data from the patient with the third party, with the HCP as mediator. The ontology metadata system is to make data understandable for the patient, who needs to be informed what kind of data will eventually be shared or kept. The same applies for the third party, but for conducting research, the third party needs to fully understand the meaning of the data.

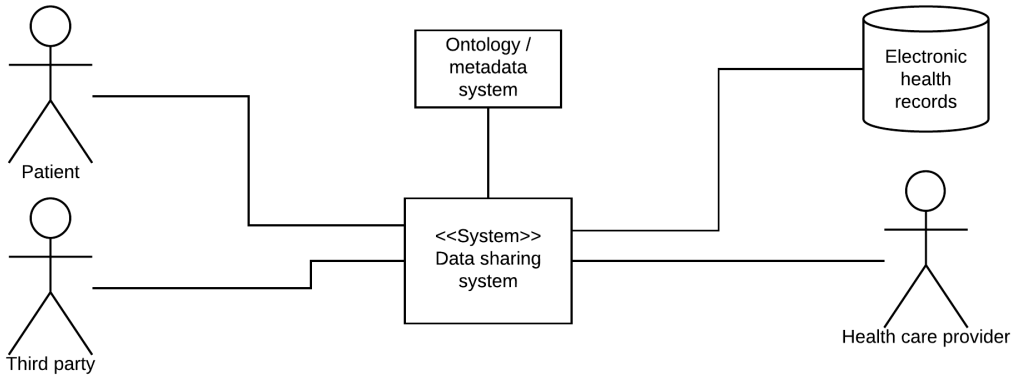

Figure 3: Context diagram of the data sharing system

4.1.2. Use case diagram

A use case diagram describes what a system must do. The diagram shows the basic functions for the different stakeholders and how they will interact with it, on a basic way. In figure 4, the proposed system is divided in tasks, represented in eclipses, the tasks are bounded to different stakeholders, represented by the stick figure. The three stakeholders are represented: the HCP, the patient and the third party.

The HCP has four tasks in not specified order. The first is to take data sources, that is done by taking the data which is stored in the EHR and give that data to the third party. The second task is to give meaning to the data. The reason to give meaning to data is that patients can understand their data, because between patient and professional, there is a knowledge gap and to bridge that gap, the data needs to be defined somehow. The third task is to contract a collaboration with a third party. The contract give boundaries to the third party. The third party must sign that contract (first task of the third party). The last task is to provide data, so the third party can use the data source for research.

The patient has one task and that is reading what the third party wants to get and for what kind of research. From that reading, the patient can give its consent. Pseudonymised data is derived from the consent of the patient. The third party can collect the data source (second task of the third party), by having only the pseudonymised data out of the consent, by having only have the anonym data or both.

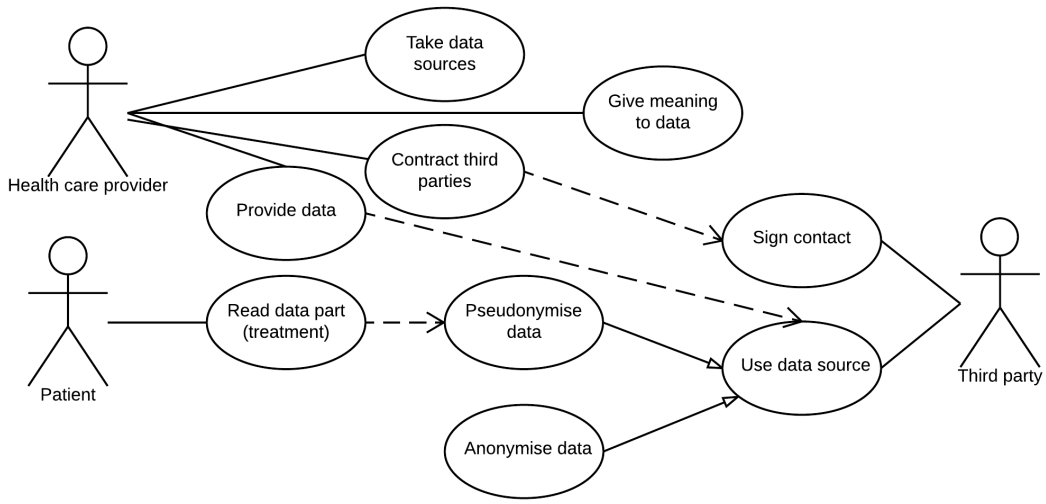

Figure 4: Use case scenarios of data sharing with a contracted third party

4.2. Functional viewpoint

In this viewpoint, the functional architecture (FA) is shown. The functions of the system and their interactions with other functions are here described. According to Rozanski and Woods, a functional viewpoint is the cornerstone of the architecture. The format of the model consists of simple boxes and arrows. Another option would be an UML based diagram, but the boxes and lines strategy seems to be more user-friendly and less strict. The

boxes represent the function spaces. The boxes can be placed in a bigger box, which represent an environment or a cluster of functions. The lines represent the interaction with other functions, with a short description on it. Lines can also come out of nothing, which stands for a first interaction with a system function.

Four clusters can be distinguished from figure 5: (1) portal, (2) data pool, (3) research environment and (4) EHR bridge. First, at the portal, the patient can manage her/his patient affairs. The portal in this figure must be understood as part of a bigger patient portal. The portal is only accessible for the patient, the data pool only for the third party. The data pool gives the pseudonymised or anonymised data to the third party. The research environment can receive a research proposal from a third party, however there ends the influence of the third party.

The EHR bridge is a bridge between the big databases of the EHR and the functions of the data sharing platform. The bridge is a crucial part and can be compared to an ETL or staging platform in the data warehouse architecture. The HCP can of course manage all the four clusters. Table 6 elaborates the functions in detail.

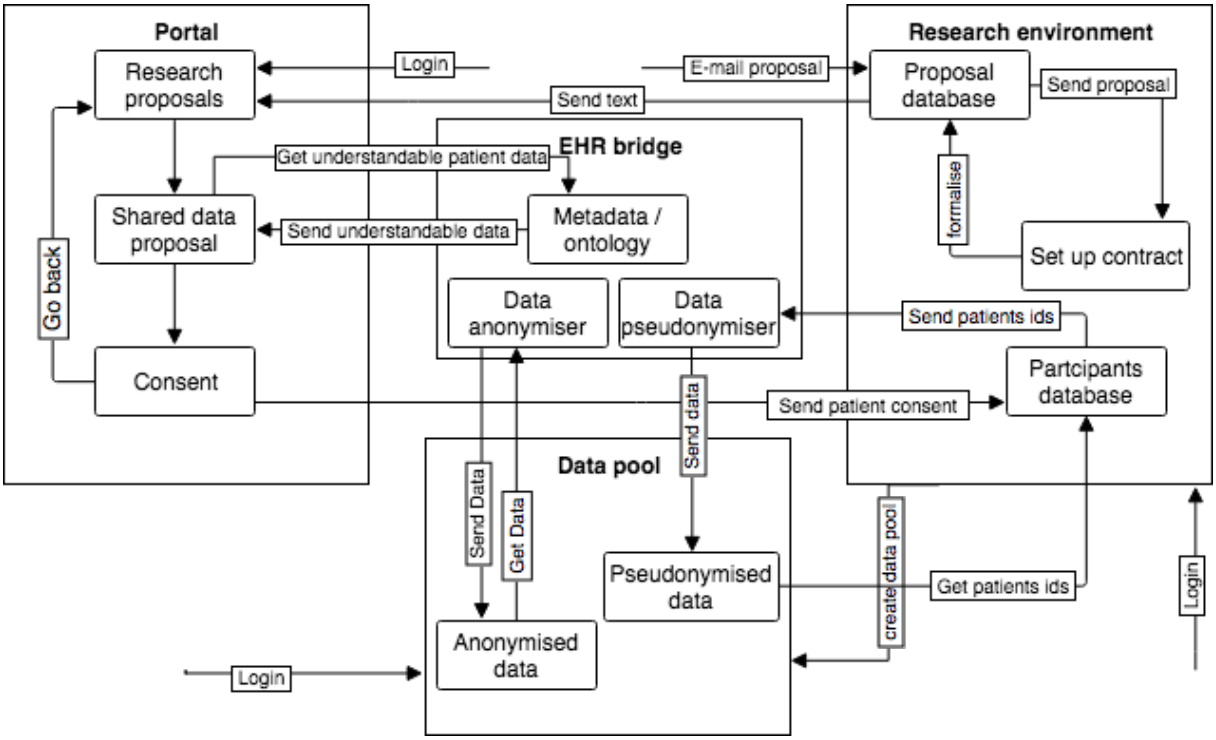

Figure 5: FA of the data sharing system

Cluster	Function	Description

(1) Portal	Research proposals	The research proposals part is where the patient can see all the current proposals and where the patient can give consent to share its own data for research.
	Shared data proposal	If the patient clicks on one of the proposals, the patient can understand what it would share.
	Consent	The patient can give its consent for a data sharing proposal. The patient's action will be administered in the participant database.
(2) Data pool	Anonymised data	The anonymised data function contains anonymised data from the EHR. The data is derived from the data anonymiser, where the EHR data is anonymised. In this place, the third party can access the data. Several solutions exist for letting the third party use this data: The third party can simply download them through a secured TLS connection and work with it on its computer or use a VPN connection to work directly on the computers of the HCP. This is all part of the security perspective in section 4.2.1.
	Pseudonymised data	The pseudonymised data function stores pseudonymised data derived from the data pseudonymiser. This data can have more details. Pseudonymised data is data where the patient gave their consent for to use it. About the connection, the same applies from the anonymised data as for the pseudonymised data and will be further elaborated in section 4.2.1.

(3) Research environment	Proposal database	The proposal database stores and saves the research proposals from the third party. The proposal will be formalised through a contract between HCP and third party. The database also connects to the portal of the patient.
	Set up contract	The set-up contract is a function to create a contract and send this contract to the third party. This way, the third party is bounded to certain rules (explained in section 4.3).
	Participants database	The participant's database stores all the patients who signed or declined cooperation with a third party. The database is connected to the EHR bridge.
(4) EHR bridge	Metadata/ontology	The metadata/ontology function is the part where the EHR will be transformed into understandable data. This data will be transmitted to the portal.
	Data anonymiser	The data anonymiser will anonymise the selected data set from the EHR. What important is, is to have some sort of threshold, such as a high k-anonymity or to use distorted data with differential privacy techniques. What should be taken into mind, is the data should not be published publicly if there is a high utility on the data set.
	Data pseudonymiser	The data pseudonymiser is a bit different than the data anonymiser. This function removes direct traceable values of a person, like names, exact birthdays and house numbers. The difference with the anonymiser is that the data is more valuable than the previous

		described function, but also harder to collect.

Table 6: Function description of the FA

4.2.1. Security perspective

An important part of designing an architecture is to take security into account. Activities are to identify security policy, identify sensitive resources or design the security implementation. This section does that for the function for letting the third party work with the data. This paper assumes that the security for protecting the EHR of the HCP is already in effect. However, a new entrance must be created for the data transfer. In section 3.3, several possible solutions are adduced. The PEP solution is not suitable to implement: the data is stored outside the current EHR and is thus not suitable to implement for this case. Besides, the system is not available yet. A simply USB-drive solution is also no suitable, if a patient decides to retreat from her/his consent, the update is not directly noticed by the third party. Maybe this is implementable in the first phase, but it will be devious in the end.

Three generic solutions are left: a remote environment through a VPN gateway, download the data from a server with a TLS handshake, or through a file sender service like SURFFilesender. The next three figures show where the third party can work, this is visualised by a box of data analytic tools.

Sketched in figure 6, a remote environment through a VPN gateway is shown. The model corresponds to the CBS case and works with the third party directly on the HCP environment. The advantage of this solution, is that the data will not proliferate to other environments, only the results can be downloaded. The disadvantage is the difficulty, a key token for the login must be purchased, the set up can be difficult, the internet connection must be very good and different kind of licences for tools must be purchased.

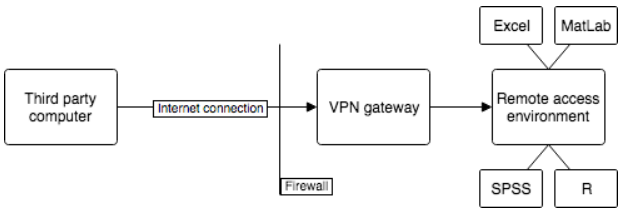

Figure 6: FA of the VPN remote environment

Figure 7 shows the download method, the third party can download the data from the server and work with it directly. With TLS handshake protocol, the third party can download the encrypted data from the FTP server. The advantage of downloading the data is that it is

simpler than a remote environment. The data is also protected if an attacker intercepts the connection. A disadvantage of this solution is that the data proliferate outside the HCP environment.

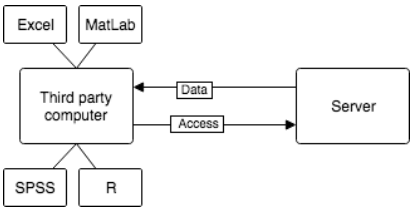

Figure 7: FA of the download method

The method through a transfer party, like the SURFFilesender, is also an option. The data is temporarily stored at a server outside the environment. This solution sounds a bit more unsafe, but this is secure through a contract with this transfer party. The same constraints as the USB-drive solution are applicable here, for every future change a new data set must be send via the transfer party to the third party. This solution can be helpful if the transfer party has an API for automatisation or some other option to automate this. The method is sketched in figure 8. Like the previous solution, a disadvantage of this solution is that the data will proliferate outside the environment.

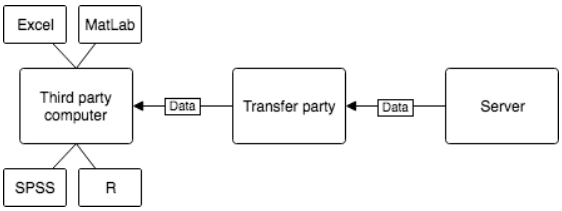

Figure 8: FA of the transfer party setting

4.2.2. Usability perspective (internationalisation perspective)

The usability perspective has a desired quality to ease the way the people can work with the system. The title of this part suspects there are two perspectives in one: indeed, in this case, usability comes close to internationalisation. The desired outcome of internationalisation is "the ability of the system to be independent from any particular language, country, or cultural group" (Rozanski & Woods, 2011, p. 597).

This part focuses on the knowledge gap between data of the patient and HCP, whereby this thesis assumes that the health care professional has more knowledge about the data than the patient does. As explained in sections 3.1.2 about consent and 3.2 about regulation, patients (or better: data owners) must be informed to give consent. To give

consent, this thesis argues, patients need to know what their data means, before they can share it. Otherwise how can a patient be informed about its actions if it does not understand what their data consists of.

Section 3.4 about defining data gives an insight into semantics and shows that through ontologies and metadata the meaning of data can be revealed. The next table shows how label and data don't say much without the meaning. With their meaning been put next to it, patients can have a better understanding of what the data means. The codes are used in different systems, such as DSM-5. Table 7 shows here that for data sharing in psychiatry existing systems can help to automatize the understanding of data. However, it is out of scope to evaluate all the different systems at different medical departments and their vocabulary.

System	**Label**	**Data**	**Meaning**
DSM-5	Diagnosis	F41.1	"Generalized anxiety disorder" (American Psychiatric Association, 2013, p. 222)
Global Assessment of Functioning (GAF), from DSM-4	GAF Score	21-30	"Behaviour is considerably influenced by delusions or hallucinations or serious impairment, in communication or judgment (e.g., sometimes incoherent, acts grossly inappropriately, suicidal preoccupation) or inability to function in almost all areas (e.g., stays in bed all day, no job, home, or friends)" (American Psychiatric Association, 1994, p. 32)

Table 7: Examples of data derived from medical systems with their meaning.

4.3. Information viewpoint

The information viewpoint "describes the way that the system stores, manipulates, manages, and distributes information" (Rozanski & Woods, 2011). The description for this section will be first written down through a process deliverable diagram (PDD), because of the combination of activity processes and the way of how information will be stored in these activities. A regulation perspective will review important parts in the PDD, which must comply with the law.

The next three sections show three PDDs. PDDs are proven to be effective in analysing and designing stages for meta-modelling. A PDD is a combination of a UML activity diagram for the activities and an UML class diagram for the deliverables. Tables give additional descriptions about processes and deliverables (Weerd & Brinkkemper, 2009).

The UML class diagrams present concepts, which can be used as a deliverable or something tangible. Some activities or concepts with a black shadow are closed concepts or closed activities, meaning that there can be more behind it than shown. The closed items also show that these sub-activities or concepts are out of the scope of this research. With a white shadow/box behind, the activity or concepts are open and will be further elaborated in the document. All activities, sub-activities and concepts are explained in the tables underneath the figures.

The information viewpoint is divided in three phases and one regulation perspective. The first phase shows the initiation phase of the data sharing process. The second phase is for letting the patient give their consent. The third phase is the stage phase and shows the process of staging the data and what will happen if the patient will change its consent. The regulation perspective connects different aspects of the system to the law.

4.3.1. Initiation phase

Figure 9 shows the main process between the health care provider and the third party. The third party can also be an internal party, when a group of scientists of the HCP wants to move data to another environment. For simplicity, third party is the prevailing name of the stakeholder.

The third party creates a proposal, with why it should have the data and what kind of data it must have from the health care provider. A research subject is helpful for communication and classification of the process. The public interest in the proposal is to comply with the in 3.2.2 described law of the complaints and disputes in health care. This law states that health care organisations must improve their services. If data sharing can improve national health care, the proposal is met in this sense. Scientific purpose, is closely related to the public interest, but not the same. This principle can help to ensure why the data is needed, for both the provider and the patient. The interest of the patient, is to let patients apply to share their pseudonymised data. The interest of the patient is in this case just as important as that of other roles and can contribute to let the patient give its consent.

Next, the HCP checks if the third party is reliable, which person from the third party is responsible for the shared data, what is the purpose is for the data, how will the data sharing take place, which data is needed and will this data be pseudonymised or anonymised. When the data is transported extern, an external security audit is also needed for a data protection impact assessment. The audit maps the risks and security in detail. A discussion with the department head and the data manager is needed to see if the research is useful, but also to check which specific data is needed. A threshold for a minimum number of participants must be determined. The criteria about the data is provided in the data elements part. The provider should bind these criteria legally in a contract and send it back to the third party. The third party then has the choice to sign the contract, when the third party is intern, the third party does not have to sign the internal regulation.

The HCP should set up the project and will ask, through the patient portal for the pseudonymised data and otherwise collect the anonymised data. These collected data will eventually be put into a data pool.

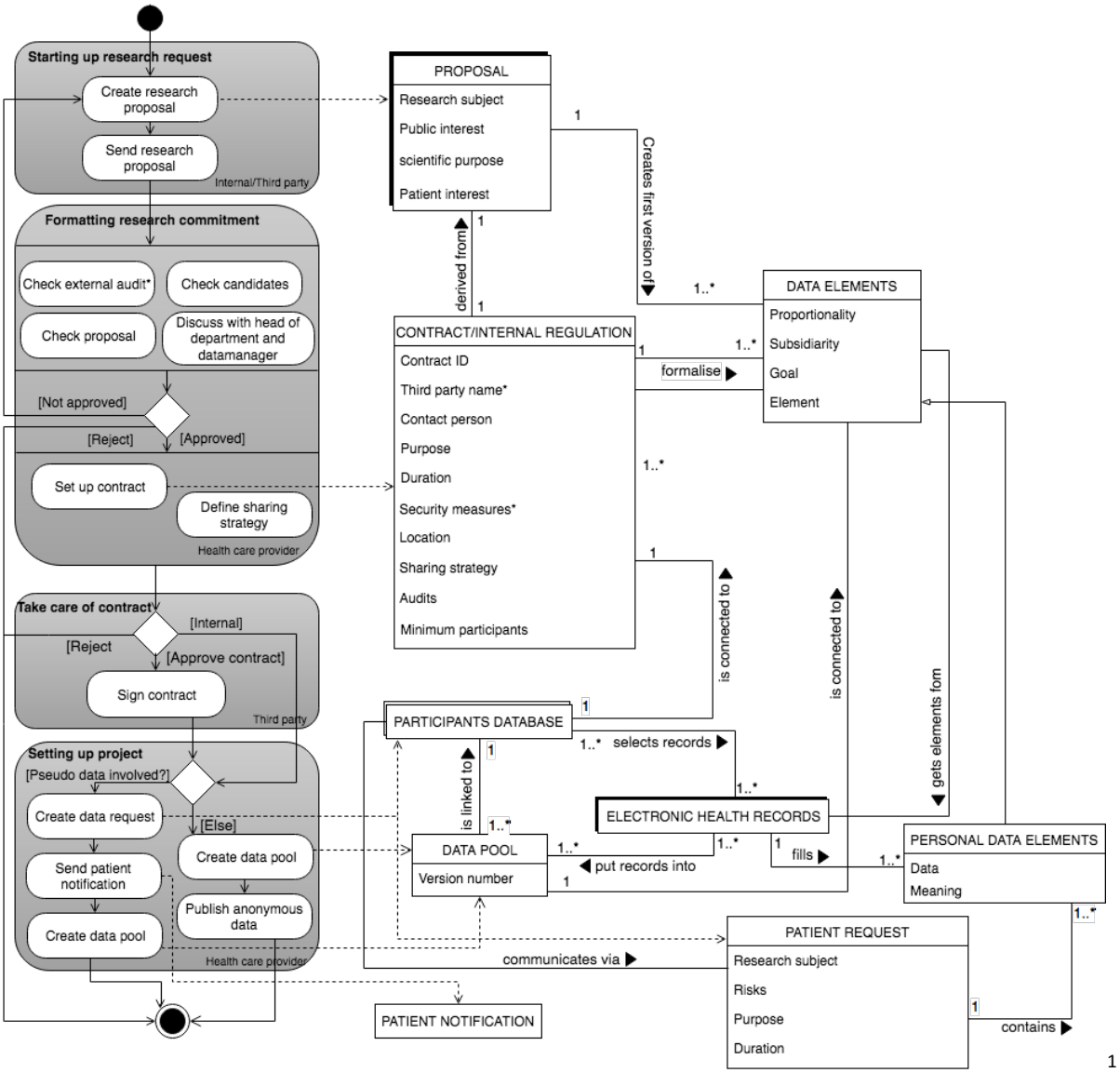

Figure 9: PDD of the initialisation phase

Activity	Sub-activity	Description
Starting up research request	Create research proposal	A research PROPOSAL is created. This PROPOSAL is further elaborated in table 9.
	Send research proposal	The PROPOSAL with the DATA ELEMENTS must be send to the HCP. The exact way how this is sent, is not

* The asterisk stands for that in some situations this activity or attribute of a concept is not needed. See the tables for more information.

		important.
Formatting research commitment	Check proposal	The HCP checks the PROPOSAL and the DATA ELEMENTS. This check is based on all the attributes, but also if the party is trustworthy, well-known and if they are commercial or public. The provider can send the PROPOSAL back and reject it, ask for additions or approve it.
	Check external security audit	This point is added after the interview with the information security officer. An external security check necessary at the third party. The process checks how the data will be saved and who can access it. Also, a data protection impact assessment will be carried out, because the data is very sensitive (EU, 2017). This step is not necessary, when there is a VPN solution to let the third party work with the data.
	Check candidates	If the target group is very small, then the privacy of the individual will be endangered. The candidates check is to check if there are enough candidates to participate.
	Discuss with head of department and data manager	Another feature is to have a discussion with the head of the department and the data manager, to see if the research is worthy and which data is needed. This discussion is needed to see which data is available and if there are other ways to do not use that data. The reason to have a discussion with the head of a department and a data manager, is because the head of a department can know if the proposal is scientific valid. The reason for the data manager is because it knows which data is saved.
	Set up contract	Based upon the PROPOSAL, the CONTRACT will be created. In the CONTRACT there are also one or more DATA ELEMENTS, which are also provided by the

		PROPOSAL. The HCP must formalise these DATA ELEMENTS for the CONTRACT. The CONTRACT states what the third party can do with the data, but also what is forbidden.
	Define sharing strategy	Together with the setup of the contract, the provider must decide which sharing strategy (pseudonymised or anonymised) is needed for the project. Potential risk, linkage of the data, although regulated in the contract, must also be weighed.
Take care of contract	Sign contract	The third party has the choice to sign the contract or retreat. Further negotiations are not modelled, but can always occur.
Setting up project	Create data request	If pseudonymised data is requested, the HCP can create a data request. Patients can now see the new research project in their health care portal. The HCP should link the metadata to the data of the patient and fill in the PATIENT REQUEST. Also, a PARTICIPANTS DATABASE must be created, to put potential candidates in.
	Send patient notification	The patient notification will send in a discrete way (without further information about the research) that there is a new notification in the patient portal. This can be done by e-mail or letter.
	Create data pool	A new data pool must be created, where the third party can work in or just download the provided data.

Table 8: Activities of the initiation phase

Concept	**Description**
PROPOSAL	The PROPOSAL is a closed concept. The third party is responsible for it, but four attributes are necessary for success: research subject, scientific purpose, public interest and patient interest. Thereby the PROPOSAL contains one or more DATA ELEMENTS, these elements

	describe what kind of data is needed for this research.
CONTRACT / INTERNAL REGULATION	The CONTRACT is important and binds the third party and the health care provider juridical and is based on inter alia processing agreements (Berg, 2017). This concept includes ten attributes: (1) contract ID, (2) third party name*, (3) contact person, (4) purpose, (5) duration, (6) security measures, (7) location, (8) sharing strategy, (9) audits and (10) minimum participants. The contract ID is for identification. The third party name is only necessary when the third party is not part of the HCP. The contact person can make a bridge between third party and the HCP. The purpose is useful to communicate with the patients and is derived from the PROPOSAL. Duration is for limiting the collaboration. Security measures are bounded regulation how the security is arranged (this is also being done in the check external security audit activity). Location is where the data is stored (can be intern, but also extern). Sharing strategy defines if the data is anonym or pseudonymised. Audits elaborate that the HCP has the right to audit the external environment. The minimum participants attribute is the threshold for giving out a data set. The CONTRACT has one or more DATA ELEMENTS and elaborated which of the data be exchanged and why. The third party should already take care of the biggest part, but the HCP will check everything, discuss it with those with authority about this data (head of department and data manager) and formalise the DATA ELEMENTS. The CONTRACT can also be an INTERNAL REGULATION, when the request comes from inside the organisation. In that case, the INTERNAL REGULATION does nog need to be signed. Third party name is an element with an asterisk, which means it is not necessary when the data is used for an internal department. The same applies for security measures, when the data stays within the department.

DATA ELEMENTS	The DATA ELEMENTS are initiated at the PROPOSAL and formalised in the CONTRACT. The DATA ELEMENTS contain four features, (1) element, (2) proportionality, (3) subsidiarity and (4) goal. The element is derived from the EHR and states which data element it is. When this is filled in by a third party, the third party doesn't have to know exactly which data elements there are, but the third party must have some clue. Proportionality is the description of the infringement of the data element relative to the importance. Subsidiarity is the report that there are no other ways than to infringe this data element. The goal is the description of how this infringement will serve the scientific and societal purpose. This part will later be formalised when it is a part of the contract by the HCP.
PERSONAL DATA ELEMENTS	The PERSONAL DATA ELEMENTS part shows the patient which of its data can be shared and for what usage. Besides the data, the meaning of the data is helpful for a patient for understanding. This deliverable is an inheritance of DATA ELEMENTS and contains all the elements of its parent.
PARTICIPANT DATABASE	The PARTICIPANT DATABASE tracks if patients have given their consent or not. The PARTICIPANT DATABASE is always connected to a CONTRACT. The database can also be linked to a DATA POOL for the pseudonymised data. The database here is an open concept and will be further elaborated in table 11.
ELECTRONIC HEALTH RECORDS	In the EHR, all the records of the patients are saved. Out of this, pseudonymised or anonymised data is derived into a DATA POOL.
DATA POOL	In the DATA POOL the third party can find their requested data. There are two hypothetical ways to work with the data. The first way is to download the data through a secured connection (such as via the TLS protocol or with a file transfer party). The second way is by working on a virtual computer in the environment of the provider. More about connections in section 4.2.1.
PATIENT	The PATIENT REQUEST is a part of the patient portal and contains four

REQUEST	features, namely (1) research subject, (2) risks, (3) purpose and (4) duration. The PATIENT REQUEST also has the personalised DATA ELEMENTS. The research subject is just for letting the patient know what the request is about. The patient can read the purpose, the risks (also derived from the external security audit) and the duration of the research. The patient can always retreat from participation, however if the researcher already got results from the data set, the retreat of that specific data is not possible anymore.
PATIENT NOTIFICATION	The PATIENT NOTIFICATION is just a way to inform the patient to go to the patient portal. The communication form of this notification can be by e-mail or letter. Important to note is that there must be no information in the notification about the research subject, department which treated the patient or other privacy sensitive details.

Table 9: Concept description of initiation phase

4.3.2. Continuous consent phase

Figure 10 shows the process of the patient in the data sharing process. Table 10 and 11 give detailed information about the different activities and concepts of the model. The patient goes for a treatment to the hospital/HCP, where the treatment is recorded in the EHR. In this analysis, the treatment is a black box and can take in any health care department. In the patient portal, the patient can go to the research proposals and see what kind of data requests are available. The patient can read information about the different data requests and what kind of data they would like to have for research. How the patient reads information is partly described in section 4.2.2. To give consent, a patient just signs a checkbox with an 'I agree to give consent' text and click on the submit button. The patient always can retreat their consent. The patient can declare if it wants to be informed about the research outcomes and if it wants to be informed about possible personal suspicious findings.

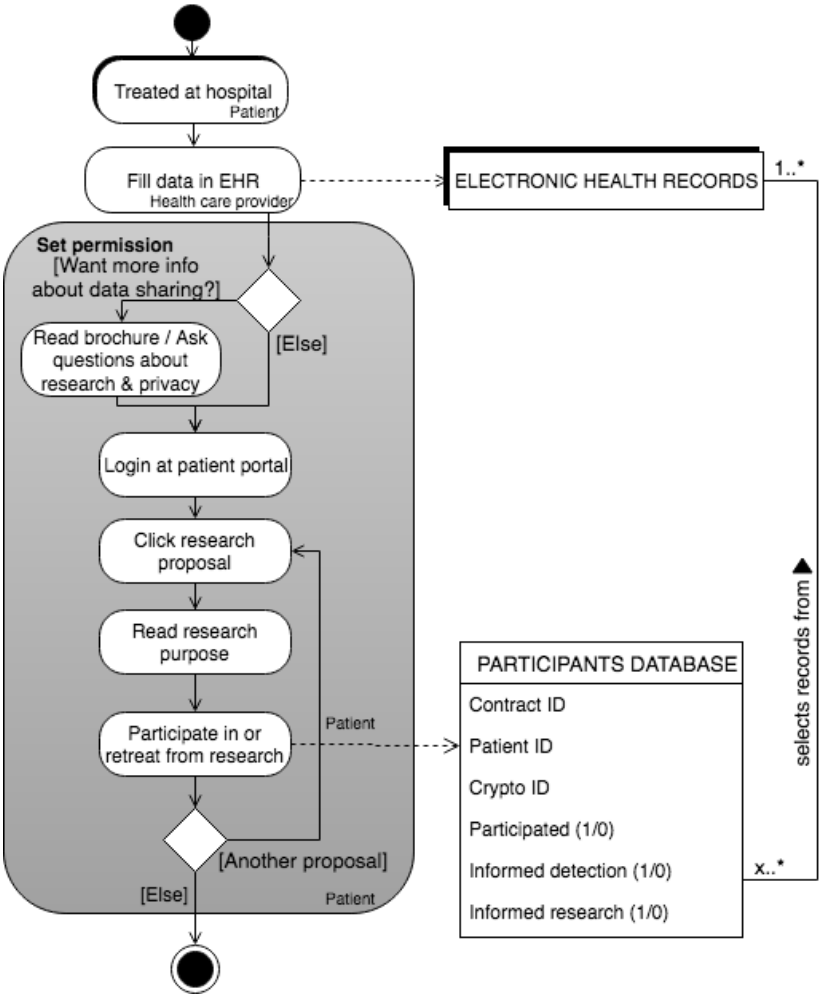

Figure 10: PDD of the continuous consent phase

Activity	**Sub-activity**	**Description**
Treated at hospital		Patient goes to the health care provider for a treatment.
Fill data in EHR		Health care professional fills the EHR.
Set permission	Read brochure / Ask questions about privacy and data sharing for research	Patient can choose to read information about scientific projects and safety and privacy and ask questions these subjects at the HCP.
	Login at UMCU portal	Patient logs into the portal and can check bills, appointments, but also participate with its data in research.
	Click research proposal	After clicking a research proposal, the

		purpose is shown.
	Read research purpose	Here the patient reads about the purpose of the project. The patient will also find what kind of data will be shared, for how long the data will be shared and what kind of data will not be shared. Thereby, the patient must also be informed that retreating of its consent is possible.
	Participate in or retreat from research	The patient now will be fully informed and can give her/his consent or retreat from the research project. The patient can go back to other treatments or research proposals.

Table 10: Activities of continuous consent phase

Concept	**Description**
ELECTRONIC HEALTH RECORDS	The professionals at the health care provider fill the EHR of the patient. It is a closed concept and it's bounded to an environment which lies out of the scope of this process.
PARTICIPANTS DATABASE	The PARTICIPANTS DATABASE has six attributes, a contract ID, patient ID, crypto ID, participated (1/0), informed detection (1/0) and informed research (1/0). The contract ID is for the connection with the CONTRACT. Patient ID is for connecting the EHR. Crypto ID is in combination with the informed detection for if the researcher finds something interesting about the patient, the patient can be informed. The patient must on forehand activate this option through informed detection. The informed research is for the patient to be informed about the results of the research. The participated attribute is to see if the patient participated in the request.

Table 11: Concept description of continuous consent phase

4.3.3. Continuous stage phase

Figure 11 shows how the pseudonymised data is staged, table 12 and 13 give a detailed description about the activities and the concepts of figure 11. The model begins with the patient, who can change its consent about a certain research. It is not modelled in the figure, but it can be interesting to ask 'why?', when the patient changes or gives a negative consent about a data sharing proposal. The reason for not modelling this, is to have a clean impression of the processes. In the administration model in section 4.5.1 takes care of these logs.

The next phase is for the HCP: If data sharing is in the production phase, the staging procedure begins. The current data pool needs to be updated. First this is done by getting data out of the EHR, then the data needs to be pseudonymised. Then after the new data set is created, the process of removing old data follows.

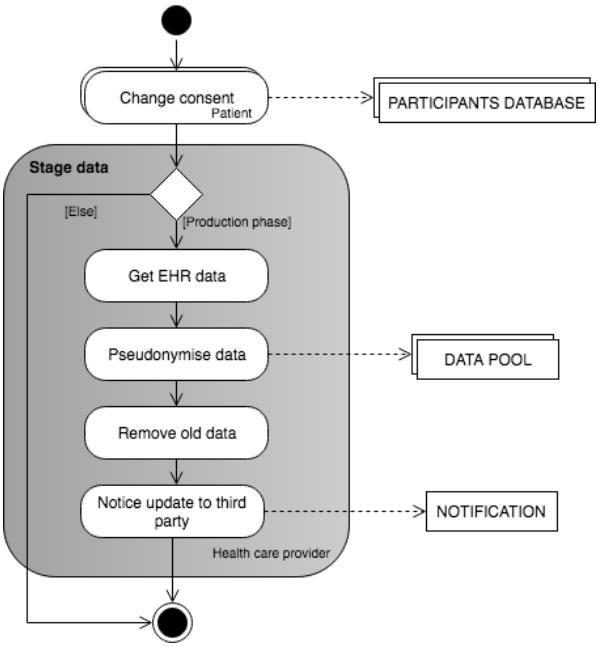

Figure 11: PDD of the continuous stage phase

Activity	Sub-activity	Description
Change consent		The patient changes its consent about a certain research. This is an open activity and is described in figure 9.
Stage data	Update pseudonymised	If the research is in production (which means there are enough participants) then this activity creates a new

	data pool	environment for the data. A new directory is added with a new version number.
	Get EHR data	The patient ID's in the PARTICIPANTS DATABASE with a consent are taken. With the contract ID, the corresponding DATA ELEMENTS are selected. With the combination of both, the data will have selected from the EHR.
	Pseudonymise data	In this activity, the data from the previous activity will be adequate transformed (based on the CONTRACT and DATA ELEMENTS) and placed in the DATA POOL.
	Remove old data	The old data at the data pool will be removed from the environment.
	Notice update to third party	The third party needs to be noticed, when the data set is changed. A simple e-mail is sufficient to complete this task.

Table 12: Activities of the continuous stage phase

Concept	**Description**
PARTICIPANTS DATABASE	PARTICIPANTS DATABASE is elaborated in table 11 and is the same concept.
DATA POOL	DATA POOL is elaborated in table 9 and is the same concept.
NOTIFICATION	The NOTIFICATION is just a simple e-mail to notice the third party that the data set needs to be updated. Important is that the old data set need to be destroyed. If the third party already has results out of the old data set, the research doesn't have to be overdo.

Table 13: Concept description of continuous stage phase

4.3.4. Regulation perspective

The regulation perspective is an important perspective in this information viewpoint, because it is needed to comply with the European and Dutch regulations. This regulation perspective explains what points in the activities and concepts of the previous figures are necessary to comply with regulations. A general description about these laws are stated in

section 3.2. In the next table, we mention different parts of the above-mentioned models to point out why they are necessary. In many cases, referring to the GDPR. The law gives rules to apply to, sometimes on very specific manners. The Dutch law comes in place for the simple reason that when data is shared without consent, the physician–patient privilege is broken.

Parts	**Measure**
Read research purpose activity (see figure 9)	With this activity, the patient can give a full informed consent. This is a first step to apply to the GDPR, that makes a consent a clear affirmative act. In here the patient must know the organisation and what is the purpose of the research. The more the patient knows, the better.
Change consent activity (see figure 11)	This activity gives the patient the opportunity to give or retreat their consent. The patient should always have the possibility to withdraw their consent. This is written down in Article 7.3 of the GDPR (EU 2016/679, 2016).
Participate in or retreat from research activity (see figure 10)	The patient must be fully informed after the previous steps. An empty ticking box that must be marked can be sufficient to comply with the GDPR.
Stage data activity group (see figure 11)	This action is needed in the production phase. If the patient withdraws from its participation, then it needs to be removed from the shared data. The reason for the removal is because the patient is owner of the data (Rumbold & Pierscionek, 2017).

Table 14: Regulation pointers to the PDDs

4.4. Deployment viewpoint

The deployment viewpoint “describes the environment into which the system will be deployed and the dependencies that the system has on elements of it” (Rozanski & Woods, 2011). The dependencies come to life through technology dependency models. Different parts of the system will be divided into smaller concepts with suggestions to bring that smaller part of the system into running. The used model looks like a functional architecture

and shows different functions of the proposed architecture in relation with existing systems, like for example DigiD (governmental secure login) and Hix (EHR software)

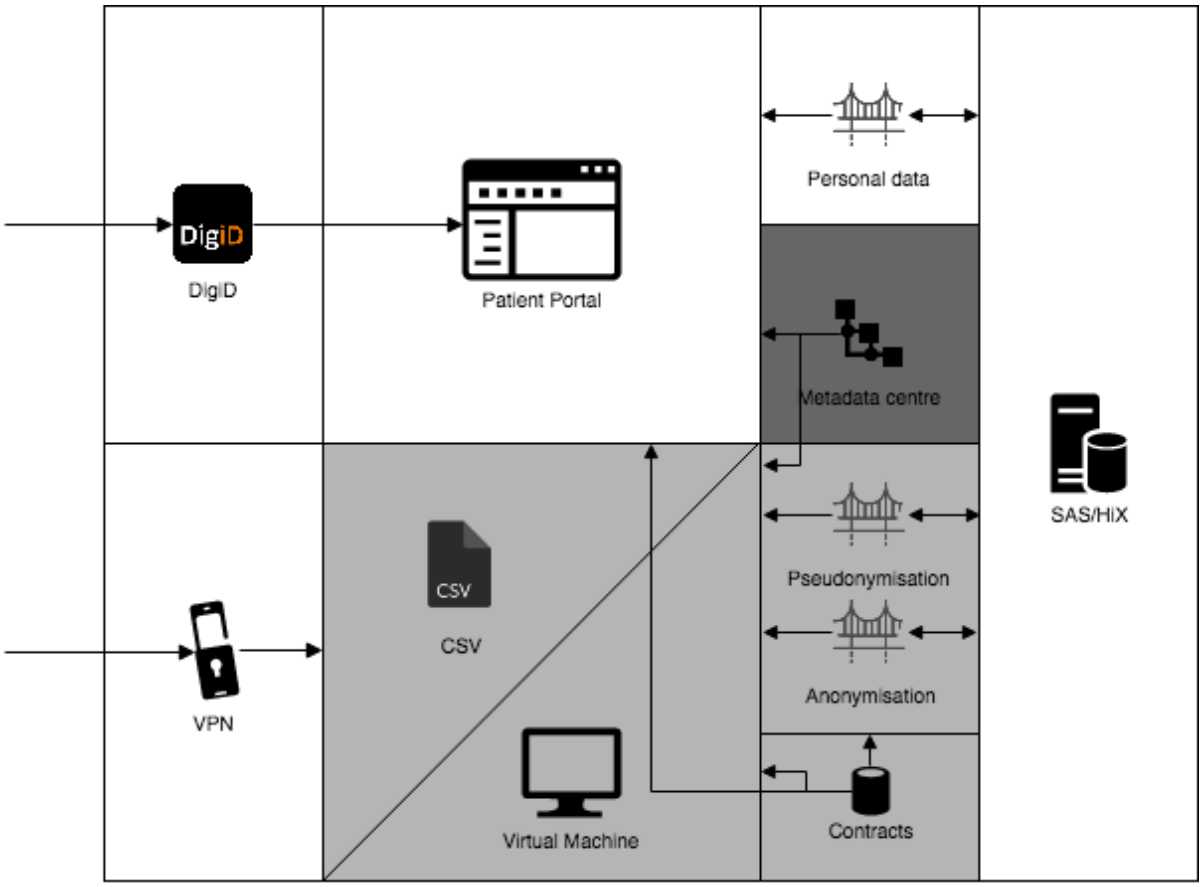

Figure 12: Deployment diagram

Figure 12 shows the deployment of the data share system in the current UMCU environment. The white areas show what is already created and for which only some changes should be made. The light grey areas need to be implemented and there are no current systems which do a similar task. The darker grey area is possibly operational, but does not behave as is should be. The next table elaborates the different systems.

System	**Description**
DigiD	DigiD is an operational system which lets patients to log in to all kind of governmental or sub-governmental websites. The system is placed outside the environment, but it authenticates the patients for the patient portal and thus is bound to one person

Patient Portal	At the patient portal of the UMCU, patients can find their dossier with treatment reports, treatment goals, patient letters, medicines, agenda, e-consulting, patient records and various other information. For data sharing, a new part must be created to show which parties want to see different kind of data of the patient.
Personal data	The personal data is a bridge between the patient portal and HiX/SAS and function as a connection for loading the patient dossier. The bridge is already in operation for data collection to see treatments in the patient portal, it only needs to be extended to connect to a data sharing request.
Metadata centre	Metadata centre is coloured in a darker grey area. Currently, there is already information about the treatments and medicines of a patient. On the other hand, not everything is specified and so patients can ask their treating physician for more information.
VPN	Nowadays, employees can log in from an external place at the UMCU environment through a VPN connection. There is a two-factor authentication. The first authentication step is with a normal username and password. The second step is with a RSA SecureID token generator, which gives an extra generated code to log in. Such a VPN connection is also suitable for working with or downloading the shared data.
CSV/FTP	The CSV/FTP part is the place where the third party can download one or more CSV-files. A simple FTP server is all what is needed to let the third party download the CSV. The CSV file is a suitable format and can be converted to different file types.
Virtual machine	The virtual machine is an advanced way to share data with the third party. As mentioned before, the great benefit is the non-proliferation of data because data is saved in the HCP environment. The environment can be installed with common data tools such as R, SPSS and Excel.
Pseudonymisation	The pseudonymisation part is a bridge between HiX/SAS and the CSV/FTP part. The bridge should remove names, identification numbers and other unnecessary items of a data set. Human authorisation is still

	desirable.
Anonymisation	The anonymisation part is a bridge between HiX/Sas and the CSV/FTP part. The bridge removes parts of the data set, like the names above, identification numbers and other unnecessary items. To comply with anonymity measurements, such as a 15-anonymity (in the k-anonymity measurement), it must do more. The bridge must aggregate data, such as birth dates to age ranges or postal code to postal areas. Another option is to add noise in the data, for example if the bridge is working according the differential privacy method. As mentioned in section 3.1.1, thresholds for anonymity in data sets can be 'hacked', via public and other (possible stolen) data sets, individuals can be linked out of an anonymous set. On the other hand, we also argued that these data sets can also be excluded from open publication through the form of contracts and eventually be destroyed after a while. If this latter is the case, we suggest that a 15 k-anonymity is sufficient. The third party must do everything to not leak the data set, since it is very easy to link leaked data sets to other data sets.
Contract	The contract part is a database which coordinates the three bridges and the CSV/FTP part. Important information is stored here, such as which forms of data sharing are involved (pseudo or anonym data), data sets, contact persons. A relational database management system like MSSQL, MySQL, Oracle database is sufficient for this task.
HiX/SAS	The HiX/SAS part is the EHR of the UMCU. Extracting data is possible, as we noticed at the data science team of the psychiatric department. This team extracts data from the HiX system for analysis.

Table 15: Deployment specification

4.5. Operational viewpoint

The operational viewpoint "describes how the system will be operated, administered, and supported when it is running in its production environment" (Rozanski & Woods, 2011, p. 393). Rozanski and Woods give different models as a suggestion for this viewpoint, such as installation, migration, configuration, administration and support models. The

administration model is the most suitable model for this case, because when the system is running, certain processes need to be monitored. Installation and migration are irrelevant, because for installation this system relies on other systems which are already in operation and this system does not need to be migrated. Configuration is elaborated at the deployment viewpoint. Support is useful if the system is already running.

4.5.1. Administration model

Administration models can be helpful to monitor and control facilities of the system. Based on the PDD, described in section 4.3, an administration model will be presented in the next section. The objective for this model is to see where the system must be monitored and how to administer the processes to detect flaws and stagnation.

Figure 13 shows a simplified version of the models of section 4.3, with the three stakeholders. The squared boxes show generalised activities, which hold much more sub-activities. The eclipses are important register logs for monitoring activity. In these registers, all activity is saved. For example, if a patient logs in the system, a new record is created. The patient number is hashed for privacy, because the logs are only meant for monitoring and to see flaws, therefore the patient number is not necessarily needed. IP is helpful to monitor where the requests are coming from, although the IP is not a sacred form of traceability (because of TOR-networks, proxy's, etc.). IP ranges can be blocked if a lot of illegal request are coming from a certain area. The browser can be saved to see on what devices or browsers patients are working, this is purely for usability. Time and data are necessary to see when a patient is doing something. This is also helpful after a data request campaign has happened, to see what the response is for a certain communication medium. Page shows simply what action the patient is doing at which page. Table 16 gives examples about how those logs looks like and are inspired on Apache webserver logs.

"Patient:c4ca4238a0b923820dcc509a6f75849b ip:99.99.99.99 time:19.35 date:2017/02/14 page:index browser:Mozilla/5.0
Patient:c4ca4238a0b923820dcc509a6f75849b ip:99.99.99.99 time:19.42 date:2017/02/14 page:datasharingindex browser:Mozilla/5.0".

Table 16: Example of logging in the patient portal

In table 17, the registers will be further elaborated. The activities will not be elaborated, because they are like the activities at the PDD at the information viewpoint. Not all activities need to be automatically monitored, because the frequency of actions is low and can be checked in a manual way. The start research data request and do research activity are not necessary to monitor, because it is situated at the third party. Formalisation data request activity and Setting up data request activity are not necessary to monitor, because the activity is done internally by the HCP and it is in line with expectations that there will not be that many requests for data sharing.

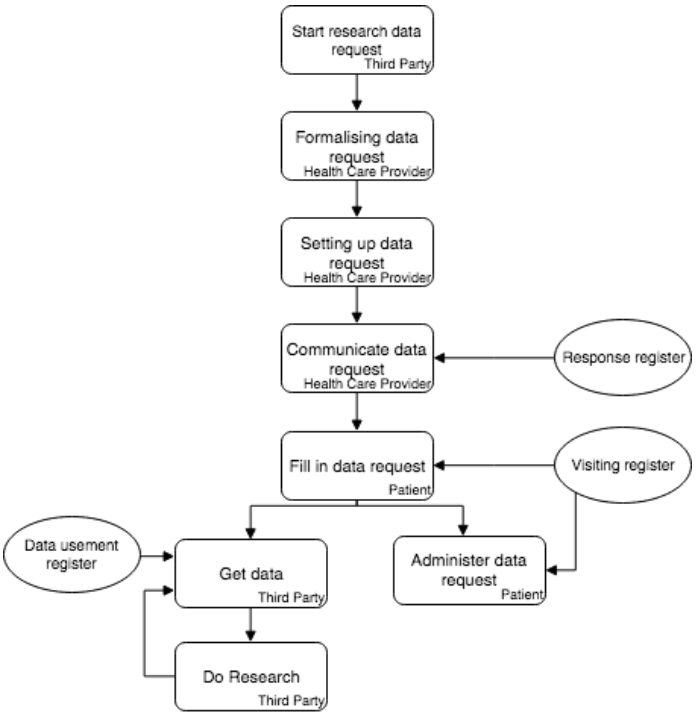

Figure 13: Administration model, eclipses give position of important logs

Registers	**Description**
Response register	The response register is just a log to know if a campaign is successful. With programs like Mailchimp, where it is easy to send thousands of e-mails, the response can be monitored. Of course, with paper letters it is difficult to monitor the response. Thereby it is helpful to check to whom a message is send out.
Visiting register	The visiting register is to monitor the patients at the patient portal. The register gives an insight into their behaviour and is

	interesting to see if they click for consent, not to give consent or do nothing after reading a research purpose. With this information, purposes can be reformulated to get more consent for data sharing. It can also be helpful for best practices in the future for these requests.
Data use register	The data use register logs when the third party downloads or works with the data. This is needed to check when a patient changes her/his consent, the third party updates her/his dataset. The log registers when the data set is downloaded, from which IP, from which third party and when.

Table 17: Register description table.

5. Validation, main contribution and limitations

This section begins with the validation of this thesis. The second section explains the main contributions. The third section gives an overview of the limitations of this work.

5.1. Validation

Rozanski and Woods (2011) provide different ways of validating the system architecture. In this study, expert reviews are used to check for limitations, redundancy, opportunities, missing items and completeness of the architecture. The proposed architecture is validated with five different persons, who are experts in their own domain. With the feedback from these experts, the architecture is reviewed from different perspectives and additions are made in the architecture to improve the quality. Table x shows the experts with the validation methods.

Expert	**Validation**
Data science leader	Walkthrough session
Patient	Walkthrough session
Information security officer	Review session
Information manager	Review session
Security officer	Review session

Table 18: Experts and validation methods

5.1.1. Walkthrough session

The walkthrough with the data science leader and a patient was to judge if the activities made sense and were 'logic' to follow. The walkthrough followed the steps of the information viewpoint. The activities and the concepts were produced on paper as far that was possible. Fake data sets were used to simulate the patient's records. With a data sharing research proposal, the attributes of the proposal had to be checked. Then, based on the proposal, a contract had to be constructed. Then the contract was signed and the wanted data was enriched by meta data. The turn was now for the patient, who could read the research proposal, check the data and sign the consent. The data of the patient is then provided for the third party. In a later stadium, the patient resigned the consent and a new data set is published.

Two obstacles were found in this session. The first obstacle was the check of the proposal. The check had to be more complex to align it more to internal responsibilities of different functions. First the storage of data is very complex, millions of records are saved, there are impassable tables. So, in the session we found out that there must be a form of consultation between the head of the department, the data manager and the one who deals with the incoming proposals. The second obstacle was the contract, the participants judged that this was not necessary when the proposal came from within the department.

The conclusion was that the tasks for the health care provider could get more specified. This because data within the EHR is very difficult to understand and therefor a collaboration between different professionals within the organisation is needed to initialise the setup between two organisations for data sharing. Positive feedback was that the patient said that if she would receive a diagnose without meaning she never would be signed. The exact meaning of data is therefore very important for the informed consent.

Action	**Name**	**Description**
Add activity	Discuss with head of department and datamanager	A discussion with the data manager and head of the department was added for assigning which data was needed. The reason for this is that the collection of data in a health care organisation is enormous.
Add	Check candidates	Check candidates to see if there are enough candidates

activity		for the proposal.
Change concept	CONTRACT/INTERNAL REGULATION	CONTRACT was changed to CONTRACT/INTERNAL REGULATION. This was for better understanding that the procedure could also be used intern.
Add activity	Send patient notification	Send patient notification activity was added for letting the patients know about the data sharing proposal.
Add concept	PATIENT NOTIFICATION	A concept which belongs to the precious activity.

Table 19: Changelog walkthrough

5.1.2. Information Security Officer

The first review session was with an information security officer. The focus in this session was on how to secure the data outside the HCP environment. The officer already had experience with other data share initiatives for large studies along with other academic health care centres. The three main additions for the architecture were to ensure there is an audit session at the other organisation when data transfers to another environment, a way back for researcher to the patient (for when the researcher found something interesting about the patient and the patient wanted to be informed) and the alertness to the potential risks of data moving out of the environment.

A point of discussion was on how to transfer the data exactly. In section 4.2.1, three methods are described to let the third party work with the data. The first is with a remote desktop inside the environment and was preferred. But in the case when the data should be transported to the CBS, this is not possible and then the method with a file transfer party in the middle is favoured above the download method. Throughout the interactive session, three questions were asked. Under the questions, the summarised answers are written down:

- What are the possible risks of data sharing in the presented architecture?
    - One of the biggest risks is the proliferation of the data. Every time the data will move from the environment, a risk of a data breach can occur. For transport therefore, you always need an encryption.

  - For the third party, you need a processing agreement, with a degree of security, the right to audit the external environment, what to do with data breaches and the capture of a time limit to report the data breach.
  - The HCP always stays responsible for the data, although the data is stored elsewhere.
- Are there any redundancy parts in the architecture? If yes, which?
  - A file transfer is not necessary when the data won't move outside the environment.
- Is there something missing in the architecture?
  - A way back of turning back from researcher to physician. When a researcher finds something interesting in a record of a patient, the researcher should be able to inform the physician of the patient. If the patient signed for informing when there is something found interesting, the patient can be helped by its physician.

Action	**Name**	**Description**
Add attributes to concept	Crypto ID and Informed detection	At the PARTICIPANTS DATABASE the attributes of crypto ID and informed detection was added, because the patient must be informed when something about the patient is founded, but only if the patient had chosen the option.

Table 20: Changelog after information security officer

5.1.3. Information Manager

The second review session was with an information manager, specialised in data quality, information management and information systems. This session focussed on how the process of data sharing runs. After this session, the data elements as part of the proposal, contract and data request was added to the PDDs. Redundant activities were removed (like login in, logout, for the sake of readability) and some activities were replaced. Throughout the interactive session, three questions were asked. Under the questions, the summarised answers are written down:

- What are the possible risks of data sharing in the presented architecture?

    - Proliferation of the data, even if you use a trusted transfer party. If the terms and conditions are right, there is no difference between a trusted transfer party and just downloading the data from the server.
- Are there any redundancy parts in the architecture? If yes, which?
    - Some of the activities are unnecessary and a reader can be sure that these activities are in there without mentioning it, like login and logout in a patient portal.
- Is there something missing in the architecture?
    - Proportionality and subsidiarity for data elements in the contract about the data.

Action	Name	Description
Add concept	DATA ELEMENTS	DATA ELEMENTS concept was added to comply the proportionality and subsidiarity suggestions.
Add concept	PERSONAL DATA ELEMENTS	As a product of DATA ELEMENTS, the PERSONAL DATA ELEMENTS was also added, to inform the patient.

Table 21: Changelog walkthrough

5.1.4. Security Officer

One of the last sessions was with a security officer. This session had all the contributions of the previous sessions processed and therefore was little to note.

Interesting in the session was that there was no distinction in the different downloading techniques. The VPN method was also favourite in this case, but for the cases of transferring the data to the third party, it did not matter according to the security officer if it was done with a transfer party or just with save downloading. About the metadata, the security officer was positive and stated: “It is a good way to give data meaning”. A contribution in the session was to give time between putting data in the data pool and consent. In that case, you give people a few days’ time to rethink their consent. Throughout the interactive session, three questions were asked. Under the questions, the summarised answers are written down:

- What are the possible risks of data sharing in the presented architecture?
    - The contract must be well closed. Add continuous auditing at the external party explicitly to the contract.

- Are there any redundancy parts in the architecture? If yes, which?
    - Not found. It looks complete.
- Is there something missing in the architecture?
    - The contract could be extended by looking at standard processing agreements.

Action	Name	Description
Add attributes to concept	Standard processing attributes to CONTRACT/INTERNAL REGULATION	The CONTRACT is extended by adding location, security measures, audits. This extension shows that

Table 22: Changelog after security officer

5.2. Main contribution

The main contribution is to give an overview of how data sharing will work in practice. The literature gives a plain state of the art of what aspects need to be considered when data sharing in a health care situation is desired. This state of the art consists law, data sharing practices such as anonymisation, transport and data understanding. The architecture following the literature study combines these aspects to a coherent solution. With different viewpoints, a framework comes to life on paper. The architecture tries to give a route map for how to start up such a data sharing system with modelled artefacts, without being dogmatic of the implementation.

5.3. Limitations

With the main contributions in our minds, this study also contains limitations. The big limitation of this research, is that it is not implemented. In that sense, this paper can be recognised as a big thought experiment on paper, based on the current state of the art and experts from for example law. This thesis provides enough artefacts to consider developing the system.

Second limitation is the external validity, although two of the experts came from other organisations, this thesis is still a case study at one organisation. For future research, it is recommended to test this architecture at multiple academic health care organisations.

This applies to the functional, information and the operational viewpoint. The deployment viewpoint is disregarded, because this viewpoint is only created for the UMCU case.

Because the system is not created, it is not possible to test the system in full operation. This is a limitation of the current study, although a walkthrough is being done. In future research, different aspects should be tested: how patients will react on a data sharing request; to what extent is anonymity necessary when it is bound in a contract (in context between the k-anonymity family till differential privacy); how to create a safe virtual machine environment for the third party.

The next limitation is part of the main contribution, in the previous part it is namely stated that it gives a route map without being dogmatic. This paper leaves some open issues of how to develop and implement for example the anonymity and pseudonymisation bridges. There are not many ready-made solutions for these concepts. Harvard is working on a data anonymisation tool for anonymous data sharing between researchers, but unfortunately this is not suitable for a health care environment. In this case, the advice is to let a development team create the bridges between the EHR and the data pool environment.

6. Conclusion

This thesis tries to answer the following question: “How, in a psychiatric health care organisation, can data sources with privacy concerns be shared with other like-wise organisations?” The question is torn apart into seven sub questions. Five of them are answered in the literature section and will be evaluated here in short. The latter two are answered without noticing in the results section and will be elaborated here in more detail. At last, this section answers the main research question.

The first question: “Which privacy regulations apply for this case study?” We see that in the Netherlands four laws are applicable. One European and three national laws. We learn from all of them that data sharing without consent is only possible if there is anonymisation, otherwise HCP must ask for consent, which must be informed. The latter remark, also answers the second question: “How to give consent?”

The third question, “How can we protect the privacy of the clients?” looks like the first question, but there is a subtle difference. This question is answered with the anonymisation techniques, like the k-anonymity family and the differential privacy solution.

But there is more. The concluding was to add contracts with the third party when anonymisation is involved. What can and can't the third party do with the data sets.

About data transfer, this thesis asked: "How can we transfer data from one place to another?" Three possible solutions are given. One of them was not operational, that was the PEP-framework. The other two showed implementable solutions. The first was just downloading data sets through a secured TLS handshake connection. The other was with a VPN connection working on a virtual machine. This thesis prefers the latter, because of the non-proliferation of data sets.

The next questions, focus on the patients. The first is: "How can patients understand their data?" A tricky question, because pure understanding is hard to measure. This is due to several factors. Like how intelligent is the patient and is the patient's understanding the same as the physician. But those ambiguous factors were not considered. The focus was on the terminology and what data means. The meaning of data was mostly projected with metadata and ontology libraries, like DSM-5, which can give a detailed insight into what data means. In table 15, a practical insight was shown that presumably the ontology sets are not sufficient. In patient portal, there is already a lot of data explained for the patient about their data, but some results are not specified and if they want to know what it is, patients must ask their physician for clarification.

The other patient minded sub question is: "How can clients choose what to share with whom?" This question is answered through many sections of the results. This is showed in the functional, information and deployment viewpoint. The system leans heavily on the already available patient portal, with an addition to let the patient choose which parties they want to share their data with.

The last sub question is answered in the deployment model: "How can we implement the new data sharing workflow into the current?" With the deployment diagram of figure 12, the facets of the system are shown in the current situation of the UMCU. This thesis believes that the best solution is to use the already created aspects of the environment, such as the VPN environment. In this environment, employees can login with a token key and is very useful for the third party to work on a virtual machine.

Back to the main question: "How, in a psychiatric health care organisation, can data sources with privacy concerns be shared with other like-wise organisations?" This question is answered by providing a data sharing system, modelled in this thesis. This study focused on

patients, since the implementation of the law must not be neglected. Patients must be informed fully to give thorough consent. The other focus was on data transfer, from HCP to third party. For this, different solutions can be used, but the best way is to keep as much data inside the HCP environment as possible.

7. Literature

Alderson, P., & Goodey, C. (1998). Theories in health care and research: Theories of consent. *British Medical Journal, 317*(7168), 1313-1315.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders, fourth edition.* Arlington: Washington, DC.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders, fifth edition.* Arlington, VA: American Psychiatric Association.

Autoriteit Persoonsgegevens. (2017). *Gebruik van medische gegevens*. Retrieved 2017, from Autoriteit Persoonsgegevens: https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/gezondheid/gebruik-van-medische-gegevens

Berg, M. v. (2017). *Bewerkersovereenkomst - verwerkersovereenkomst*. Retrieved 11 9, 2017, from Justitia, objectieve en praktische juridische hulp: http://www.justitia.nl/privacy/bewerkersovereenkomst.html

Bhargavan, K., Fournet, C., Kohlweiss, M., Pironti, A., Strub, P. Y., & Zanella-Béguelin, S. (2014). Proving the TLS handshake secure (as it is). *International Cryptology Conference*, 235-255.

Chang, X., & Terpenny, J. (2009). Ontology-based data integration and decision support for product e-design. *Robotics and Computer-Integrated Manufacturing, 25*(6), 863-870.

Chassang, G. (2017). The impact of the EU general data protection regulation on scientific research. *ecancer, 11*(709), 1-12.

Christen, P. (2012). *Data matching: concepts and techniques for record linkage, entity resolution, and duplicate detection.* New York: Springer Science & Business Media.

Christen, P., Vatsalan, D., & Verykios, V. S. (2014). Challenges for privacy preservation in data integration. *Journal of Data and Information Quality (JDIQ), 5*(1-2), 4.

Dataverse. (2016, 12 06). *Private data Sharing Interface*. Retrieved 08 03, 2017, from Private data Sharing Interface: https://beta.dataverse.org/custom/DifferentialPrivacyPrototype/

De Silva, S., Liu, A., & Nabarro, L. L. (2017). Europe's tough new law on biometrics. *Biometric Technology Today*(2), 5-7.

Dijk, J., Choenni, S., Leertouwer, E., Spruit, M., & Brinkkemper, S. (2013). A Data Space System for the Criminal Justice Chain. *On the Move to Meaningful Internet System* (pp. 755-763). Springer Berlin Heidelberg: OTM Confederated International Conferences.

Doerr, M. (2003). The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. *AI magazine, 24*(3), 75-92.

Dustdar, S., Pichler, R., Savenkov, V., & Truong, H. L. (2012). Quality-aware service-oriented data integration: requirements, state of the art and open challenges. *ACM SIGMOD Record, 41*(1), 11-19.

EU 2016/679. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council*. Retrieved 04 02, 2017, from Official Journal of the European Union: http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L:2016:119:FULL&from=EN

EU. (2017, 10 04). *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679*. Retrieved 11 1, 2017, from Article 29 data protection working party: http://ec.europa.eu/newsroom/document.cfm?doc_id=47711

EU GDPR Portal. (2017, 08 09). *GDPR Portal: Site overview*. Retrieved from Eugdpr.org: http://www.eugdpr.org/

Fortineau, V., Paviot, T., & Lamouri, S. (2013). Improving the interoperability of industrial information systems with description logic-based models: The state of the art. *Computers in Industry, 64*(4), 363-375.

Gambs, S., Killijian, M., & del Prado Cortez, M. (n.d.). De-anonymization attack on geolocated data. *Journal of Computer and System Sciences, 80*(8), 1597-1614.

Gardner, S. P. (2005). Ontologies and semantic data integration. *Drug discovery today, 10*(4), 1001-1007.

Ghawi, R., & Cullot, N. (2007). Database-to-ontology mapping generation for semantic interoperability. *VDBL'07 conference, VLDB Endowment ACM*, 1-8.

Gomez-Cabrero, D., Abugessaisa, I., Maier, D., Teschendorff, A., Merkenschlager, M., Gisel, A., . . . Tegnér, J. (2014). Data integration in the era of omics: current and future challenges. *BMC systems biology, 8*(2), 1-10.

Green, M. (2016, 06 15). *What is Differential Privacy?* Retrieved 07 02, 2017, from A Few Thoughts on Cryptographic Engineering: https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/

Hastings, J., Smith, B., Ceusters, W., Jensen, M., & Mulligan, K. (2012). Representing mental functioning: Ontologies for mental health and disease. *ICBO 2012: 3rd International Conference on Biomedical Ontology*.

Heiler, S. (1995). Semantic interoperability. *ACM Computing Surveys, 27*(2), 271-273.

Hevner, A. R. (2007). A three cycle view of design science research. *Scandinavian journal of information systems, 19*(2), 1-6.

Hoytema van Konijnenburg, E., Teeuw, ,. A., & Ploem, M. C. (2015). Data research on child abuse and neglect without informed consent? Balancing interests under Dutch law. *European journal of pediatrics, 174*(12), 1573-1578.

IDABC. (2004). *European interoperability framework for pan-European egovernment services.* Retrieved from IDABC: http://ec.europa.eu/idabc/servlets/Docd552.pdf?id=19529

Jongejan, W. J. (2016, 09 04). *Psychiatrie-afdeling UMCU balanceert op randje met big-data-analyse* . Retrieved from Zorg-ICT Zorgen: https://www.zorgictzorgen.nl/psychiatrie-afdeling-umcu-balanceert-op-randje-met-big-data-analyse/

Kim, J., Lee, Y., Lee, K., Jung, T., Volokhov, D., & Yim, K. (2013). Vulnerability to Flash Controller for Secure USB Drives. *Journal of Internet Services and Information Security, 3*(3/4), 136-145.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Technical Report EBSE-2007-01, School of Computer Science and Mathematics, Keele University*.

McGuire, A., Oliver, J., Slashinski, M., Graves, J., Wang, T., Kelly, P., . . . Hilsenbeck, S. (2011). To share or not to share: A randomized trial of consent for data sharing in genome research. *Genetics in Medicine, 13*(11), 948-955.

Moreno-Conde, A., Moner, D., Da Cruz, W. D., Santos, M. R., Maldonado, J. A., Robles, M., & Kalra, D. (2015). Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. *Journal of the American Medical Informatics Association, 22*(4), 925-934.

Narayanan, A., & Shmatikov, V. (2006). How to break anonymity of the netflix prize dataset. *arXiv preprint cs/0610105*, 1-24.

Naumann, F.; Bilke, A.; Bleiholder, J.; Weis, M. (2006). Data fusion in three steps: Resolving schema, tuple, and value inconsistencies. *IEEE Data Eng. Bull., 29*(2), 21-31.

OED Online. (2017). *consent, n.* Retrieved 03 20, 2017, from Oxford English Dictionary: http://www.oed.com/view/Entry/39517?rskey=0EFPyz&result=1#eid

OED Online. (2017). *semantics, n.* Retrieved from Oxford University Press: http://www.oed.com/view/Entry/345083?redirectedFrom=semantics

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems, 24*(3), 45-77.

Polson, P. G., Lewis, C., Rieman, J., & Wharton, C. (1992). Cognitive walkthroughs: a method for theory-based evaluation of user interfaces. *International Journal of man-machine studies, 36*(5), 741-773.

PrivacyBarometer. (2016, 09 13). *Zorgverzekeraars krijgen inzage in medische dossiers bij vermoeden van fraude*. Retrieved 10 05, 2017, from Privacy Barometer: https://www.privacybarometer.nl/maatregel/100/Zorgverzekeraars_krijgen_inzage_in_medische_dossiers_bi...eden_van_fraude

PrivacySense.net. (2015, 07 09). *Different Types of Consent*. Retrieved 05 17, 2017, from Privacy information, tips and expert interviews - PrivacySense.net: http://www.privacysense.net/different-types-consent/

Rozanski, N., & Woods, E. (2011). *Software systems architecture: Working with stakeholders using viewpoints and perspectives (2nd edition).* Westford: Addison Wesley.

Rumbold, J., & Pierscionek, B. (2017). The effect of the General Data Protection Regulation on medical research. *Journal of Medical Internet Research, 19*(2), 1-6.

Sharda, R., Delen, D., & Turban, E. (2014). *Business intelligence: A managerial perspective on analytics.* Essex: Pearson.

Slater, T., Bouton, C., & Huang, E. S. (2008). Beyond data integration. *Drug discovery today, 13*(13), 584-589.

Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., . . . Lewis, S. (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology, 25*(11), 1251-1255.

Springall, D., Durumeric, Z., & Halderman, J. A. (2016). FTP: The forgotten cloud. *46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, 503-513.

SURF. (2017, 05 10). *SURFfilesender*. Retrieved 07 28, 2017, from SURF: https://www.surf.nl/diensten-en-producten/surffilesender/index.html

Ten Kate, M. (2016, 07 16). *Gedwongen opname? Op dag vijf zal de patiënt agressief zijn*. Retrieved from Het Financieele Dagblad: https://fd.nl/morgen/1160117/gedwongen-opname-op-dag-vijf-zal-de-patient-agressief-zijn

Toledo, C. v., & Spruit, M. (2016). Adopting privacy regulations in a data warehouse: A case of the anonimity versus utility dilemma. *Proceedings of the 8th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 67–72.

U.S. National Library of Medicine. (2016). *Unified Medical Language System (UMLS)*. Retrieved 09 21, 2016, from U.S. National Library of Medicine: https://www.nlm.nih.gov/research/umls/

UMC Utrecht. (2017, 08 16). *Organisatie*. Retrieved from UMC Utrecht: http://www.umcutrecht.nl/nl/Over-Ons/Organisatie

Venable, J., Pries-Heje, J., & Baskerville, R. (2012). A comprehensive framework for evaluation in design science research. *International Conference on Design Science Research in Information Systems*, 423-438.

Verheul, E., Jacobs, B., Meijer, C., Hildebrandt, M., & Ruiter, J. d. (2016). *Polymorphic encryption and pseudonymisation for personalised healthcare.* Retrieved 10 6, 2016, from Iacr.org: https://eprint.iacr.org/2016/411.pdf

Weerd, I. v., & Brinkkemper, S. (2009). Meta-modeling for situational analysis and design methods. In M. R. Syed, & S. N. Syed, *Handbook of research on modern systems analysis and design technologies and applications* (pp. 35-54). Hershey: IGI Global.

Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering. *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, 3-13.