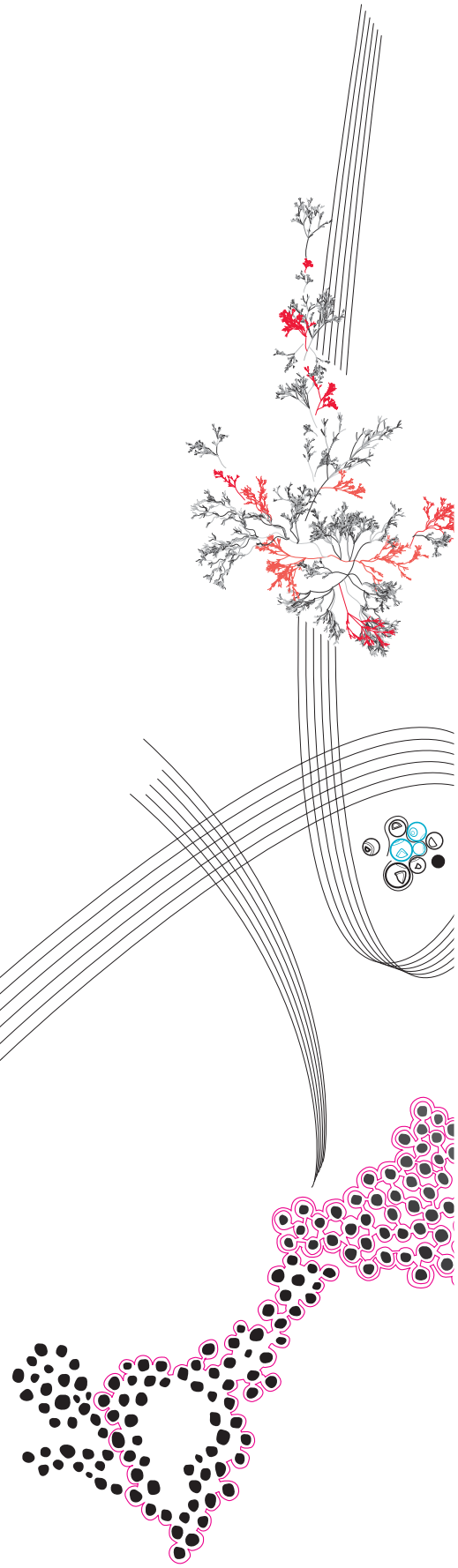


MSc Business Information Technology
Final Project

Designing a Regional Medical Dataspace: An Architecture for Healthcare Data Sharing in the Twente Region

Victor Dibbets

Supervisor: dr. J.L. Rebelo Moreira, dr. J.P.S. Piest and dr. S. Bosems

November, 2024

Department of Computer Science
Faculty of Electrical Engineering,
Mathematics and Computer Science,
University of Twente

UNIVERSITY OF TWENTE.

Contents

Abstract

This thesis addresses the challenge in the Twente region of the Netherlands surrounding ensuring interoperability between the various healthcare providers present there. The thesis presents an architecture for a medical dataspace in order to improve interoperability between healthcare providers in Twente, following the IDS-RAM. First, background literature research and stakeholder interviews in Twente produced the requirements of the dataspace, for both the primary use of medical data, as well as the secondary research use. These requirements were then utilized to produce various ArchiMate models of the theorized dataspace. These models were then validated by experts whom provided feedback based on their expertise. They provided positive points on interoperability and scalability, while raising issues about security and data quality. Overall, the architecture presented in this thesis provides a foundational approach to enhance medical data interoperability in Twente, addressing a significant regional challenge in healthcare data management using the novel dataspaces approach.

Chapter 1

Introduction

Healthcare has been a primary requirement in human life since before written records have existed. [1] In recent times however, we have also started to produce a lot of data in the patient journey. Over the years this healthcare data has increased in volume dramatically [2]. It has also been observed by several organisations that this data is not being utilized in an optimal fashion [3] [4]. One way to ensure that the data which healthcare organisations collect is used more efficiently is to enable more data sharing between several organisations in order to make the entire medical process more efficient.

Over time, a plurality of methods to share data have been developed. From sharing data to and from relational databases in 1970 [5] to the current movement for data fabrics [6] which has been gaining popularity recently; when it comes to sharing data there are many options to choose from. Consequently, many of these ways of sharing data have also found their way into medical data sharing.

One of these data sharing alternatives is that of the dataspace. This concept was introduced in 2005 [7] and was motivated by the desire to move away from traditional DataBase Management Systems (DBMS) and towards an architecture where data is easily shared but data owners keep complete control over their own data. How this is achieved will be discussed further in chapter 2. The paper further describes the relationship between traditional DBMS and DSSP (Dataspace Support Platform) as: *'Unlike a DBMS, however, a DSSP does not assume complete control over the data in the dataspace. Instead, a DSSP allows the data to be managed by the participant systems, but provides a new set of services over the aggregate of the systems, while remaining sensitive to their requirements for autonomy.'* [7].

Since the initial paper proposing the dataspace in 2005, a large number of papers have been published on the topic. Scopus [8] alone shows 660 papers which contain 'dataspace' or 'data space' in the article title. These papers cover a wide variety of sectors and topics, deepening the concept of the data space significantly. This growing body of literature underscores the potential of dataspaces in optimizing healthcare data usage, showing a potential significant advancement in the search for more efficient and autonomous data sharing practices.

1.1 Problem Statement

Data produced in the healthcare sector is also growing quickly [9]. As mentioned before, multiple organisations have noted that this data are not always used in an optimal manner.

The Twente region in the Netherlands currently does not have a regional data sharing initiative and is exploring multiple options on how to achieve efficient healthcare data sharing in the region.

One of the ways that this could be achieved is through the setup of a dataspace for medical data. However, there is currently no literature available on utilizing a dataspace for achieving comprehensive medical data availability, and commonly, a (regional medical) datahub is used to achieve this (eg. [10]). A regional medical datahub in this context means a centralized platform to collect, share and manage data within a specific geographical region, Twente in this case. The IDS-radar (a dashboard published by the International Dataspace Association on active dataspaces) [11] currently also shows no dataspaces in the healthcare sector for that purpose.

Even though it has been shown in previous publications that dataspaces show significant benefits for data sharing [12], the lack of application in a medical setting on a regional scale is a gap which is to be resolved. Filling this knowledge gap could uncover an efficient way of sharing data within a medical region, leading to more productive healthcare.

Lastly, in 2016 the paper 'FAIR Guiding Principles for scientific data management and stewardship' [13] was published. This paper provides guidelines to make data more FAIR, or Findable, Accessible, Interoperable and Reusable. When data follows the guidelines put forth in this paper, then that data will be more machine-actionable. This means that systems can more easily find, access, interoperate and reuse data without human input.

1.2 Research Objective

In this thesis the objective is to identify and analyze medical dataspace implementations in order to design an architecture for a regional datahub that satisfies the requirements of the reference architecture model for international data spaces (IDS-RAM) and FAIR data principles. This to-be designed architecture will be developed using the ArchiMate modeling language [14]. Justification for using ArchiMate will be included in section 2.7.

1.3 Research Methodology

The design science methodology (DSM) as presented by Wieringa in 'Design Science Methodology for Information Systems and Software Engineering' [15] is used for this thesis. This method involves multiple stages in order to develop an artifact. These stages are:

- **Problem Investigation:** During the problem investigation phase the goal is to identify problem areas, before the requirements are formulated and the artifact has been created. During the problem investigation stakeholder and their corresponding goals are also identified.
- **Treatment Design:** In this second phase the requirements for the artifact (the architecture) are drawn up and the artifact is designed.
- **Treatment Validation:** In this last phase of the design cycle the treatment which has been designed in the previous phase is to be validated.

Problem investigation corresponds to chapter 2 and 3. Treatment Design is included in chapter 4 and Treatment Validation in chapter 5.

1.4 Research Questions

In order to address this research objective some research questions need to be formulated. First, a main research question and stemming from this main research question, some smaller, sub-questions will also be posed. The question will be based on the design question template as described by Wieringa in [15]. The format he proposes is the following: 'How to [(re)design an artefact] that satisfies [requirements] so that [stakeholder goals can be achieved] in [problem context]¿

The main research question of this paper is:

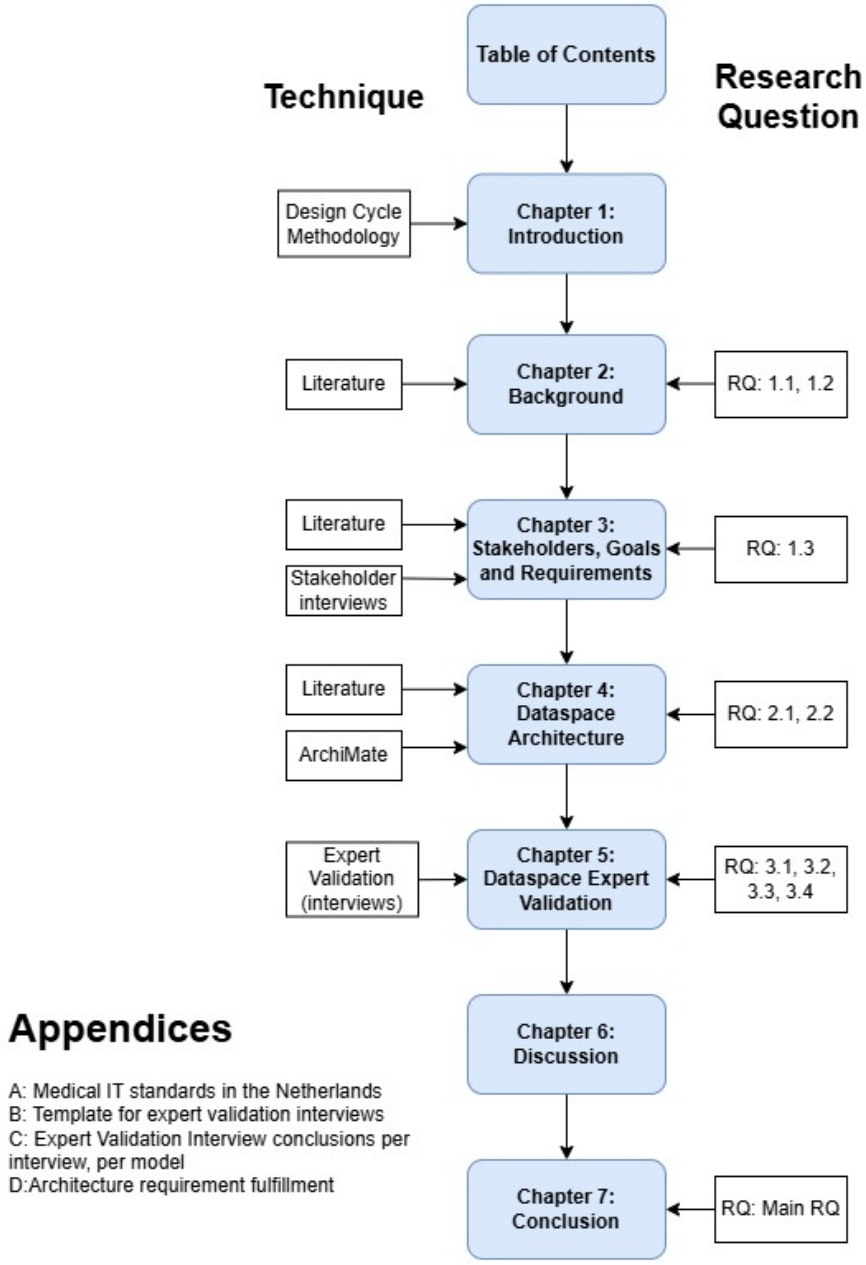
How can an architecture for a regional medical datahub be designed, that satisfies the requirements of the reference architecture model for international data spaces and FAIR data principles, to facilitate data sharing and/or exchange amongst stakeholders in the Twente region?

To facilitate the answering of this question, several (knowledge) questions are to be answered, relating to each phase of the Wieringa design cycle.

Phase	**Questions**
Problem Investigation	*1.1:* What is the current state of medical data sharing in the Netherlands? *1.2:* Which medical dataspaces are currently in operation following the IDSA-RAM? *1.3:* Who are the stakeholders for the project and what are their goals?
Treatment Design	*2.1:* What are the (non-)functional and technical requirements for a regional datahub? *2.2:* How to design an architecture for a regional medical datahub which ensures interoperability between existing standards, follows (data-)regulations, and allows for appropriate data governance policies regarding access, quality, and compliance using the IDS-RAM?
Treatment Validation	*3.1:* How do stakeholders perceive the problems identified during the problem investigation? *3.2:* What are the metrics for evaluating appropriate data exchange for a regional medical dataspace? *3.3:* Following an analysis of the architecture, could the architecture be an appropriate solution for medical data sharing in the Twente region? *3.4:* How does the architecture compare to existing data exchange solutions?

TABLE 1.1: Overview of Research Phases and Questions

FIGURE 1.1: Report structure with corresponding RQs and techniques used



1.5 Report Structure

In this diagram (figure 1.1) above you will find the entire structure of the report. For each chapter there is also an indication of what research questions will be answered in a given chapter, and what techniques are used in the writing of the chapter.

Chapter 2

Background

This section introduces required background information in order to correctly frame and introduce the dataspace which is be presented in section 4. First, dataspaces as a whole is explained. Following that, the current state of medical data sharing in the Netherlands is discussed. Third, data governance as present in the IDS-RAM is presented, followed by matters of data privacy and compliance. Next, the FAIR data principles within the context of the IDS-RAM are analyzed. And lastly, the TOGAF framework is introduced alongside the ArchiMate modeling language.

2.1 Dataspaces

In order to develop an architecture for a dataspace it is of crucial importance to understand the context of what a dataspace entails. As mentioned in the introduction to this thesis, the term dataspace was first introduced in 2005 by Franklin [7]. Since then new papers are constantly being published (Scopus listed 45 papers with the term 'Dataspaces' in the title, abstract or keywords since 2020). The aforementioned first paper on dataspaces was published by Franklin, Halevy and Maier and the goal of that paper was to propose a solution for existing DBMS not coping well with the concept of "data everywhere". The solution they proposed is that of a dataspace. The definition they provide in their paper is as follows: *'Dataspaces are not a data integration approach; rather, they are more of a data co-existence approach. The goal of dataspace support is to provide base functionality over all data sources, regardless of how integrated they are. For example, a [DataSpace Support Platform] (DSSP) can provide keyword search over all of its data sources, similar to that provided by existing desktop search systems. When more sophisticated operations are required, such as relational-style queries, data mining, or monitoring over certain sources, then additional effort can be applied to more closely integrate those sources in an incremental, "pay-as-you-go" fashion.'* As this definition states, a dataspace is not an architecture, it is a approach for data to co-exist, while still allowing for at least some (semantic) interoperability for all that data.

This definition however still does not fully answer what a dataspace is and how it differs from other database solutions or approaches. To answer this question, papers by Halevy and Dong from 2007 [16] and by Otto from 2022 [17] are helpful. These two papers add to the definition from 2005 by emphasising the (partially) unstructured data which can be present and integrated in a dataspace, that the semantic relationships between sources are not always known and that the data is not physically integrated, but rather is stored at the data source.

2.1.1 International Data Spaces Association - IDSA

The architecture presented in this paper follows the guidelines provided by the IDSA reference architecture model (IDS-RAM). Given that this is the case, it is important to know the organisation behind the architecture model. The International Dataspace Association is an organisation founded in 2017 following a project by the Fraunhofer society called 'Industrial Data Space'. This new organisation was tasked to develop a reference architecture model to assist organisations in building international dataspaces [18].

The IDSA has defined the mission and vision behind the IDS-RAM as follows: *'It is time to change the way data is shared. We want to pave the way for a data economy in which every company and every person keeps full control over their data treasures. We believe in a data economy in which you do not rely on a solution that is owned by one big player. This is why we create the required standards for data spaces which grant data sovereignty to all participants to share data without regret. Our mission is to advance the IDS standard and drive innovation, awareness and global adoption of data spaces to ensure data sovereignty, meaning staying in control of access and usage of your data, for all participants.'* [19].

The IDS-RAM is a large repository defining every aspect of an international dataspace. It is divided into four different chapters [20]:

1. **Introduction**: Introduces the goals for the IDS and the corresponding strategic requirements.
2. **Context of the IDS**: Describes the IDS in various different contexts and how an IDS can be relevant for these contexts. Sections include "Big Data and AI" or "Blockchain".
3. **Layers of the RAM**: The IDS has split the RAM into five different layers, each discussing a different aspect of the RAM. These layers are business, functional, information, process and system:
    - The **business layer** discusses various roles and interactions in the IDS.
    - The **functional layer** describes the functional requirements and features which are to be present in the finished IDS.
    - The **information layer** specifies the information model. This is crucial for ensuring that all parties which are present in the IDS are interoperable with each other.
    - The **process layer** focuses on the interactions which take place in an IDS and the processes that go along with these interactions.
    - The system layer takes the roles which are specified in the business layer and the interactions of the **process layer** and maps these onto concrete data and a service architecture.
4. **Perspectives of the RAM**: Provides three cross-sectional perspectives on the five layers discussed above. The three perspectives are Security, Certification and Governance.

2.1.2 Current IDS-RAM implementations in the healthcare sector

The IDSA maintains the IDS-radar. This is a dashboard listing a wide variety of information on all implementations of dataspaces currently in operation or development. It allows for the user to filter per sector. When this is done to only show dataspaces in the health sector seven entries remain. Below, these are each described to get an understanding of the current state of IDS-RAM implementations in the healthcare sector. The seven listed are divided into four use cases, and three dataspaces.

Dataspaces:

- **Dataspace4Health**: Dataspace4Health is a dataspace initiative in Luxembourg with the intention to 'revolutionise secure and compliant health data exchange in the country and beyond' [21]. It was launched on the 25th of March 2024 and aims to leverage data in order to create connected, secure, and efficient healthcare ecosystem that improves patient outcomes and advances medical research and innovation. It does this by following Gaia-X framework [22].

- **Flemish Health Data Space**: Very little information exsists on the Flemish Health Data Space outside of the IDS-radar. According to the IDS-radar, it aims to make the use of secondary healthcare data more efficient.

- **HEALTH-X dataLOFT**: The HEALTH-X dataLOFT is a data platform which puts a lot of emphasis on the patient being in control of their own data [23]. In this dataspace, patients can decide who has access to their medical data. Part of the data platform is that they are developing an ecosystem which is open, federated and based on the Gaia-X standards. Currently, no information is available online on how far the project has progressed so far, there are however milestones available which they work towards in the project [24].

Use cases:

- **Gatekeeper**: Gatekeeper is a data marketplace of sorts with the objective to 'connect healthcare providers, businesses, entrepreneurs, and elderly citizens and the communities they live in' [25]. According to the IDS-radar part of this project is to develop a dataspace connector extension to facilitate in data sharing between various parties.

- **Medical Data Space MedDS**: The Medical Data Space (MedDS) is a virtual space that supports secure exchange and easy integration of medical and health-related data from diverse sources, using standards and shared governance models. It aims to improve the quality of diagnostics, preventive and therapeutic measures, and the monitoring of therapies. It is a collaboration between two departments of the Fraunhofer society, ICT Technology and Life Sciences [26].

- **NL AI Coialition: Oncology Research**: This use case focuses on oncology research. The dataspace allows for data to easily be shared between hospitals in order to be used for AI-powered oncology research [11] [27].

- **VELES Project**: The VELES project is a project in several countries in southern and eastern Europe (Bulgaria, Cyprus, Greece and Romania). The project focuses on advancing predictive, preventive, and personalized medicine through health data, AI, cloud computing, and IoT. It also seeks to align regional strategies for health

data sharing, improve clinical practice, and provide secure, innovative digital health services. The aligning of regional data sharing strategies is done through a Regional Smart Health Data Space [28].

2.1.3 Summary

As can been seen above there are multiple implementations and use cases for dataspaces around the world in the healthcare. However, none of these focus on the specific scenario as described in the problem statement in section 1.1. This focus is that of a **regional** medical data hub, as a dataspace, following the **IDS-RAM** and mostly focusing on the **primary use of patient data**. Because of this, it is clear that a research gap certainly exists in this space.

2.2 Current state of medical data sharing in the Netherlands

In order to fully integrate the to-be developed dataspace into the current status quo of medical data sharing it is important to first have a full understanding of what the current status quo is in the Netherlands. This is done by looking at various aspects of medical IT systems, focusing on the standards in use in the Netherlands, current data sharing pipeline and data sharing initiatives.

2.2.1 Medical IT standards in the Netherlands

To facilitate efficient data transfer in the medical sector, many standards have been introduced. These standards range from use-case or datatype specific to broader communication standards. A wide variety of standards are described in Appendix A. In this section a deeper dive on the most common standards are described.

HL7 (FHIR): Is a range of application level standards which aim to facilitate (health) data transfer between applications. It is very widely used and has various different versions. FHIR is a HL7 standard based on RESTful operations and is also becoming increasingly more popular. Given this popularity, the focus here, and the rest of the report is mostly on the FHIR implementation of HL7. FHIR is a RESTful specification and is organized around a repository, this repository is build up out of various resources which the user can call upon. These resources are represented through a URL defined by FHIR [29]. The following is an example of a FHIR URL: `http://server.example.com/fhir/Patient/23455`. The first part (`http://server.example.com/fhir/`) is the base address, and is used to identify the server which has the information. The second part (`Patient`) identifies the service which the user wants to employ. Lastly the third part (`23455`) is the ID which manages a given instance of the resource.
FHIR specifies three base services (the second part mentioned above). These are Instance, Type and System services. An Instance service allows a client to retrieve a resource, update it to a new state or delete the resource (The RUD of CRUD). The Type service allows the user to search through existing resources or create a resource (C of CRUD). Lastly, the System service is a service which is focused on larger scale changes. The book by Benson [29] provides a diagram giving more information on FHIR resources. This can be found in figure 2.2.
A 2024 survey on the use of FHIR in various countries around the world [30] showed

that in The Netherlands HL7 is already commonly used, with an expected strong increase in the future.

DICOM: Is an international standard for transmitting, storing, retrieving, printing, processing and displaying medical images. One of the primary goals for the standard is to facilitate communication between various interested parties, both hardware and software related. According to the documentation published by DICOM themselves, it achieves this by specifying: [31]:

1. For network communications, a set of protocols to be followed by devices claiming conformance to the Standard;
2. The syntax and semantics of Commands and associated information that can be exchanged using these protocols;
3. For media communication, a set of media storage services to be followed by devices claiming conformance to the Standard, as well as a File Format and a medical directory structure to facilitate access to the images and related information stored on interchange media;
4. Information that must be supplied with an implementation for which conformance to the Standard is claimed.

A Dutch study [32] surveyed the use of DICOM in Europe and showed that DICOM is the most widely used format for image transmission and that in the surveyed countries 92% of surveyed professionals name DICOM as their preferred standard.

SNOMED CT: Is a comprehensive system of healthcare terminology. Each term has a concept code, description and can have relationships to another SNOMED CT term (e.g. Infective Pneumonia IS a Respiratory Disease). It provides the core general terminology for EHRs. SNOMED is most commonly used to ensure that the communication in EHRs is accurate and relevant and it is used in more than 80 countries. When utilizing SNOMED CT users tend to use single codes, where longer, more complex expressions are possible [29]. According to Benson, SNOMED CT has two key features, the first being that SNOMED is virtually future proof, because it can evolve. And the second being that it supports multiple parent-child relationships, which is something other coding schemes cannot.
SNOMED does not enjoy full adoption in the Netherlands due to smaller Dutch standards. However, given the recent push from the European Union for more cross-border data exchange, the international nature of SNOMED has seen it's popularity rising. Furthermore, the Dutch standards are being mapped to SNOMED terms and the aspects relevant to Dutch healthcare are also being translated [33]. At MST the terminology system which is in use is also being mapped to SNOMED CT.

OpenEHR: At its core, OpenEHR is a non-profit organisation which develops and publishes technical EHR standards sourced from the international medical informaticians community. The principal goal is to develop a lifelong, patient-centric shared health record [34]. OpenEHR is not a software one can download, however there are several tools and applications which are based upon OpenEHR that are available for use. The semantic approach of OpenEHR is based on multi level modelling, where models built by domain experts are in their own layer. These models is where data representation is defined. These domain experts can be directly involved in the development of these models, and these models can then be used to build archetypes.

Archetypes are clinical concept models (where possible content is defined) which can then be combined into templates, which is where actual content is configured [35]. Applications can then be build upon these templates which is how standardization is achieved using OpenEHR. A visualized version of this interaction can be found in figure 2.1. This is however a very brief description of the topic and for more information please refer to the white paper produced by openEHR: [35].

Table 2.1 provides a summarized version of the 4 standards above.

Standard	**Purpose**	**Standardization Organisation**
HL7 (FHIR)	Standard to facilitate electronic transfer of health data. FHIR is based on the RESTful operations	Health Level 7
DICOM	Standard for transferring of medical images	DICOM Standards Committee
OpenEHR	An open, vendor-neutral platform for managing patient EHRs	OpenEHR Foundation
SNOMED CT	A comprehensive health terminology to support healthcare semantic interoperability	SNOMED International

TABLE 2.1: Standards and their purposes

FIGURE 2.1: OpenEHR interaction between models, archetypes and templates

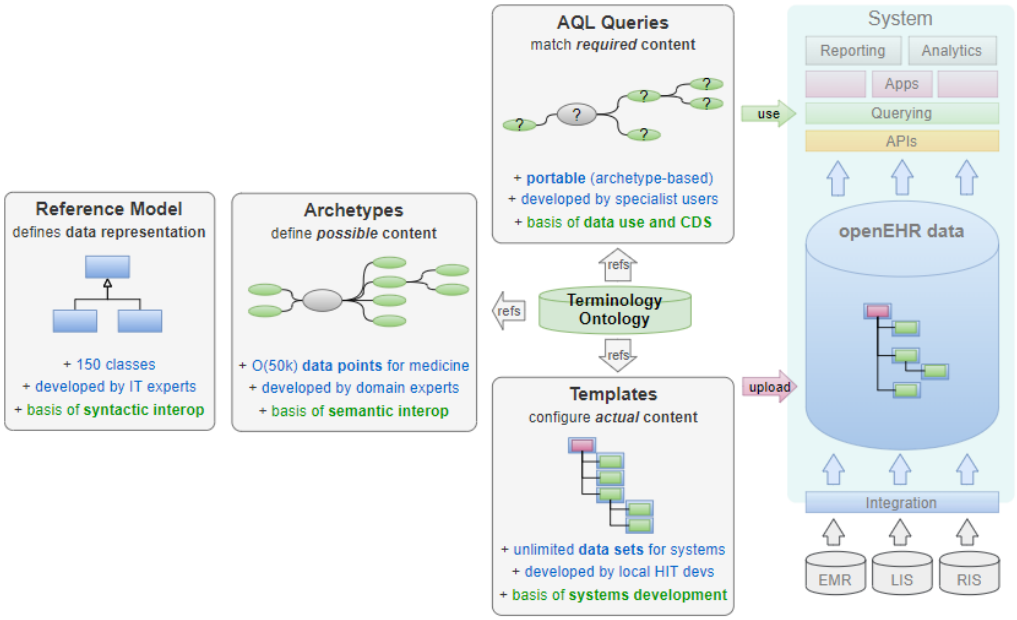
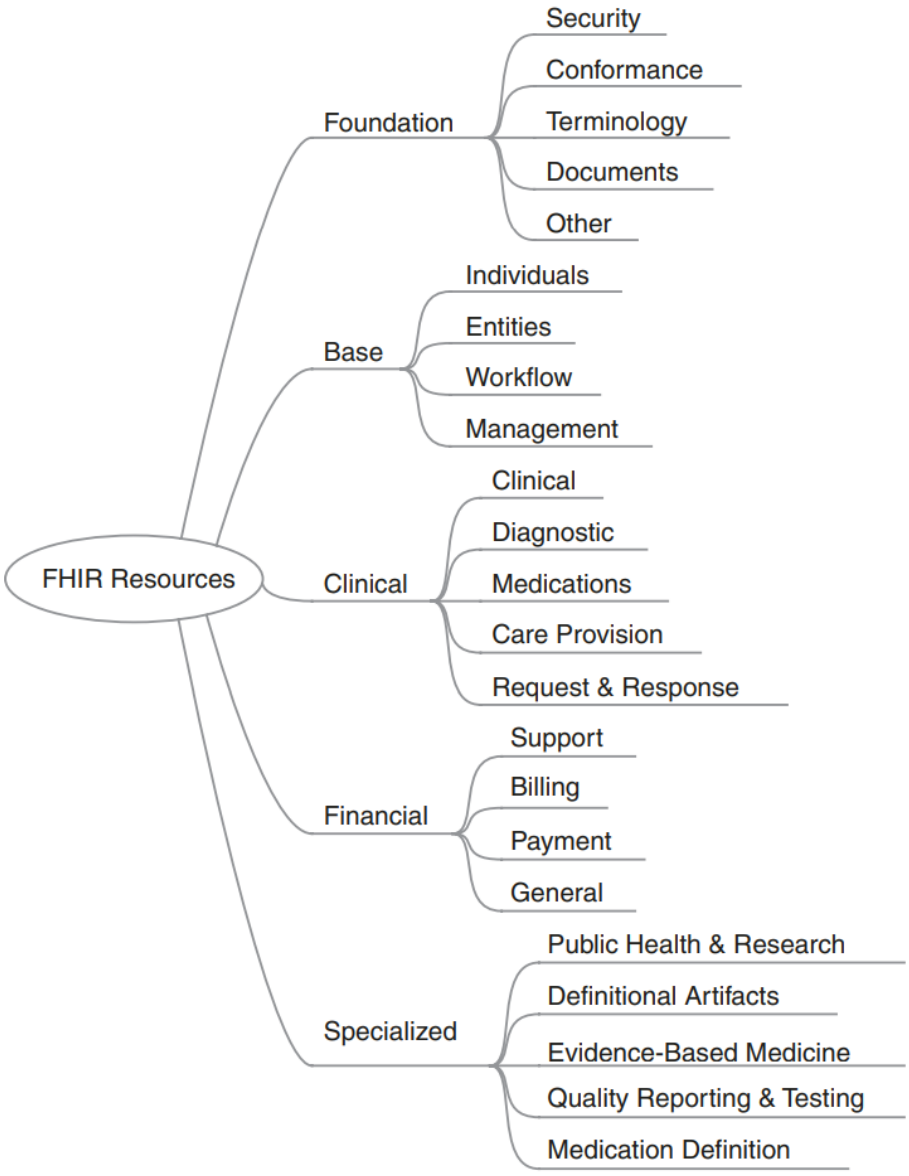


FIGURE 2.2: FHIR resources as shown in [29]



2.2.2 AORTA/LSP

One of the most common patient journeys is where patients visit the general practitioner (GP) which in turn directs them to a hospital to see a specialist. Since this journey is common there is already a well established data pipeline in place for this purpose in the Netherlands. This works through the interaction of the AORTA architecture in collaboration with the 'Landelijk Schakelpunt' or LSP (National Exchange Point). In this architecture, data (in an EHR) is always stored at the healthcare providers themselves, and not at a centralized data base. The healthcare providers can then request data on demand. This data is then provided through the LSP. The LSP does not log any medical information but it does log who requests which information in order to allow patients to audit who accessed their information.

According to the AORTA-LSP website [36] there are currently more than 4700 general practitioners offices and 71 hospitals whom make use of this system. The system has a couple of noted disadvantages. These are noted in the 'Argumentenwijzer' [37] from the Dutch Centre for Ethics and Health, which is a government initiative organisation to 'signal and provide information about new and current ethical issues in healthcare and biomedical research that are relevant to government policy.' [38]. These disadvantages include:

- Less complete medical files: When files are shared on a large scale healthcare providers could consider registering less information as they fear that their colleagues could discover mistakes or they want to truly ensure their patients privacy.
- Copying mistakes: When patient information is shared, you run the risk of one healthcare provider copying or assuming the truthfulness of a colleagues mistake. Having these copies spreading creates a form of "new reality" which could prove dangerous to patients when this data is incorrect.
- Privacy risks: Whenever data is shared you run the risk of data leaks, this is also the case in this system.

2.2.3 Standards conclusion

As you may have noticed, the standards above have many differences. This mostly stems from the fact that they all have a different purpose, within the medical domain. HL7 (FHIR) exists to ensure data transfer, DICOM to communicate image data, SNOMED CT ensures semantic interoperability, OpenEHR allows for all medical information on a patient to be stored in one interoperable, lifelong patient EHR and AORTA/LSP is for primary care data in the Netherlands. While many of these have alternatives which are in use across the Netherlands, these standards described above seem to be most prevalent based on the various sources cited above in their respective sections.

2.2.4 Data sharing initiatives in the Netherlands

In the Netherlands there are several active data sharing initiatives in the healthcare sector. Below four of these is discussed.

- **CumuluZ**: CumuluZ is an initiative with the goal of making patient data more easily accessible. They aim to do this by making all healthcare data available through one national data infrastructure [39]. The CumuluZ initiative is spearheaded by four organisations (NFU, de NVZ, mProve and Santeon) consisting of various hospitals. They currently have a basic reference architecture in place. This can be found in figure 2.3.
- **Digizorg**: Digizorg is an app developed by the Erasmus MC in Rotterdam [40]. It is an application targeted at patients in order to make the patients healthcare journey more transparent and efficient. They do this by keeping appointments and diagnoses in the app. It also features the possibility to talk to healthcare professionals at the hospital and to communicate simple measurements one can do at home like temperature or weight. Currently, only the hospital is connected to the Digizorg app but in the future they aim to add general practitioners and other healthcare providing facilities in the Rotterdam region [41].

- **Zorgviewer**: Zorgviewer is an application produced by RIVO Noord (Regionaal InformatieVOorzieningsoverleg Noord-Nederland). The goal of Zorg-viewer is to enable cross-organisational data sharing in the region. As of the 22nd of March 2024, it is not yet live. When they go live they will have doctors from 4 different organisations connected, but the aim is to increase that number over time [42]. Both Zorgviewer and Digizorg are developed in collaboration with CumuluZ.

- **HDSA**: HDSA is short for Health DataSpace Amsterdam. The aim of this project is to create a dataspace for the metropolitan region of Amsterdam. They created a regional healthcare data infrastructure where multiple healthcare institutions in the region use each others data which has been made pseudonymous. They aim to improve patient care and research in doing so. They also positioned the dataspace in such a way where it can be a step towards participation in the European Health dataspace (EHDS) [43]. The HDSA is different from the architecture presented in this thesis in a couple of different ways. First, this HDSA does not follow the IDSA reference architecture. Furthermore, the HDSA is more focused on research, rather than the patient healthcare journey.

- **Health-RI:** Health-RI is an organisation in the Netherlands which aims to improve the reuse of health data for research, policy and innovation [44]. It also serves as a national coordination point to create agreements regarding the reuse of health data. Furthermore, Health-RI also provides services to assist research in the health sector, divided into seven different "portfolios" of research. [45]. These seven portfolios are:

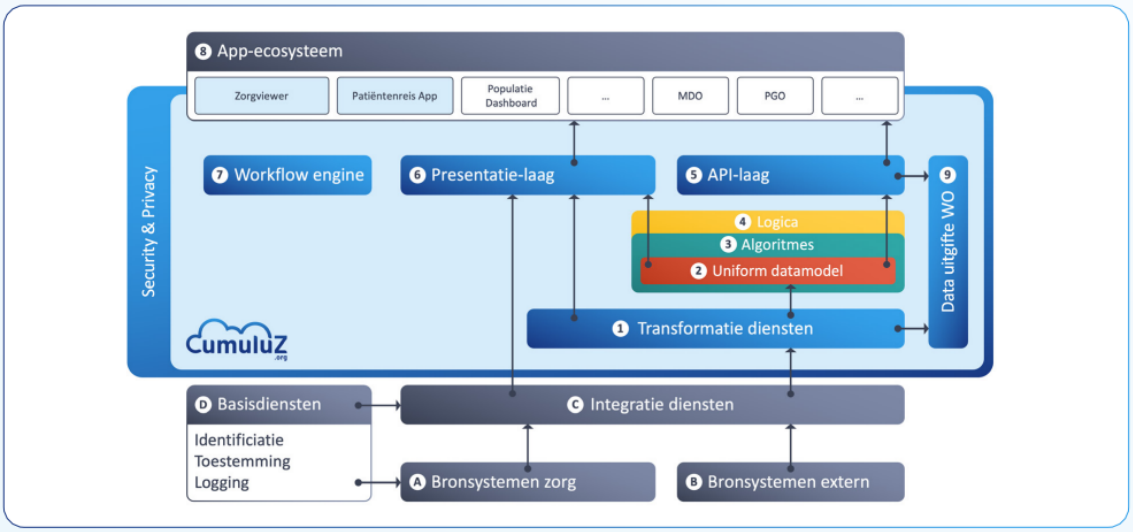
  1. Analytics & Software
  2. Education & Training
  3. Ethical & Legal
  4. FAIR Data Stewardship
  5. Health data & Samples
  6. IT & Digital Workspaces
  7. Research Facilities

Table 2.2 provides a summarized view of the discussed data sharing initiatives in the Netherlands.

Initiative	**Primary Objective**	**Region**
Cumuluz	Creating a reference architecture for a national data infrastructure	National
Digizorg	App supporting data exchange in healthcare process	Rotterdam Region
Zorgviewer	Cross-organisational data sharing	Northern Netherlands
HDSA	Improve research data sharing	Amsterdam
Health-RI	Improve the reuse of health research data	National

TABLE 2.2: Sharing Initiatives and their Objectives

FIGURE 2.3: CumuluZ reference architecture



2.2.5 Ziekenhuis Referentie Architectuur (ZiRA)

ZiRA is a reference architecture developed by Nictiz. Nictiz is a Dutch expertise center for medical standardization and eHealth [46]. ZiRA is a collection of various models targeted on how to design the organisation and information infrastructure at Dutch hospitals [47]. Given that there are multiple hospitals in Twente, and hospitals serve a central role in many healthcare processes, ZiRA provides valuable information in the development of any healthcare architecture in the Netherlands, including the one presented in this thesis. ZiRA contains a large variety of models, from business process models, to application models. In chapter 4 the exact components of ZiRA which were used for the dataspace architecture are discussed.

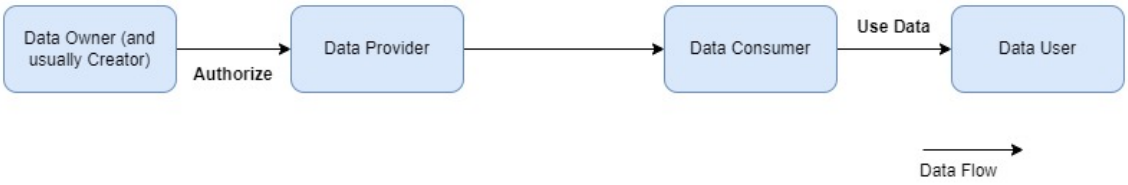
2.2.6 Summary

As discussed above, there are various aspects of medical data sharing which are relevant in the Netherlands. From various standards (both the three mentioned in section 2.2.1 but also smaller ones in Appendix A) to the initiatives which are in place in order to promote medical data sharing in the country. While these initiatives have their merits, none of them cover the research gap of deploying a regional medical data hub as a dataspace based on the IDSA reference architecture.

2.3 Data Governance in the IDS-RAM

Before applying data governance to the IDS-RAM it is first important to define what data governance is. The definition of data governance used for this thesis is retrieved from the book "Data Governance for Managers" by Lars Michael Bollweg [48]. The book defines data governance as: *'The structured integration of Data Management practices (procedures and methods) in the organizational structures and processes of a company.'*.

FIGURE 2.4: Basic data interaction within a dataspace



2.3.1 Key Roles

One important aspect of the IDS-RAM not covered in section 2.1.1 is that of the key roles and their activities in the IDS-RAM. In this section a small selection of the key roles in the IDS-RAM [49] are discussed, and what their relationship is to data governance in the system. Not all roles is discussed, but a selection is made based on what is deemed most important for this paper. A description and explanation of all roles can be found in [49].

Data Supplier: The data supplier is the role which first introduces the data into the system. The role is usually further specified into one of the following: **Data Owner**, **Data Creator** and/or **Data Provider**. The Data Owner is the entity who can execute control over the data. The owner is also responsible for defining the usage policy of their data. Usually, the data owner is also the data creator. The data provider is the entity which makes the data available for others in the IDS. By making it available the data customer can use the data for their purposes.

Data Consumer: Similarly to the data supplier role, the data customer role can also be subdivided into specific sub-roles. As the name implies, the **Data Consumer** receives the data from the data provider (ISDA calls the data consumer the mirror entity of the data provider). The consumer counter-part of the data owner is the **Data User**. The data user is the entity which has the legal right to use the data provided by the data provider, according to the usage policy of the data. The data user and consumer can be the same entity, but they do not have to be. This basic dataspace interaction between a data owner to the data user can be seen in figure 2.4.

2.3.2 Metadata

Various aspects of the IDS-RAM deal with metadata, as it is a core aspect of data exchange in an (IDS) dataspace. Metadata is defined as *'Structured, encoded data that describes characteristics of information-bearing entities (including individual objects, collections, or systems) to aid in the identification, discovery, assessment, management, and preservation of the described entities'* [50], or in simpler terms: data about data.

The IDS-RAM uses this metadata in several ways throughout the reference architecture. Primarily, metadata serves as a critical component, enabling effective data management, discovery, interoperability, and security. The rest of this section below discusses these topics in more detail, and describe how metadata contributes to the data governance in the IDS-RAM.

Data Discovery and Exchange: As mentioned above, metadata serves an important role in allowing participants in the dataspace to find the data they require. This is done in an interaction using the data broker. A data provider provides the data broker with the metadata on the datasets they are providing. The data consumer can then query that metadata to determine whether the data they are seeking is available in the data space. If this interaction is complete, then the data provider and the data consumer can start the data exchange according to the usage policy of the data provider. Additionally, metadata is also used to describe data apps which data stores publish to be used in the dataspace.

Interoperability: Metadata also serves a role in ensuring interoperability of the dataspace. First off, the IDS-RAM also states that the metadata should '*describe the syntax and serialization as well as the semantics of data sources*'. This requirement ensures that the original data source is able to be fully understood using the meta data. Second, a metadata model is also to be created for the dataspace to ensure consistency of the metadata amongst the participants.

Security: Next, there are some aspects of security in the dataspace in which metadata serves an important function. Metadata is used in the concept of data provenance. Data provenance is '*the description of the origins of a piece of data and the process by which it arrived in a database*' [51] and is used in the IDS to allow participants to see the linage of the data they provide and use. The concept is also strongly tied to the data sovereignty that participants in the dataspace have. Metadata also serves a function in other aspects of security like identity verification in the dataspace and the certification process of participants.

Other: Lastly, messages in a dataspace always have metadata as well. This metadata is provides evidence of the communication, in other words, elements which allow you to see details about sender and receiver, but also allow you to have extract context from the message, such as type of content or usage contract.

2.3.3 Identity Management and Data Access Control

One of the strategic requirements of the IDS-RAM is to aim for full trust in the dataspace. Added to this is the fact that for this thesis, the data being exchanged is medical data, which is generally speaking private data. Full trust can only be achieved through adequate identity management and data access control measures. This section discusses what the IDS offers when it comes to these topics.

Identity Management:
As the IDS-RAM is designed to facilitate secure data exchange among its participants, having appropriate identity measures in place is of crucial importance. To realize this secure exchange of data, appropriate identity verification measures must be deployed.

In this context, Identity management in IDS means the verification and authentication of each participant's identity. This ensures that only authorized participants have access to the dataspace and can exchange data within said dataspace. Identity management in the IDS works through identifiers, provided by an identity provider to a IDS connector[1].

[1]This is slightly simplified. The identity is provided to a connector with certain identity components. However, for the sake of brevity, this is not described in more technical detail.

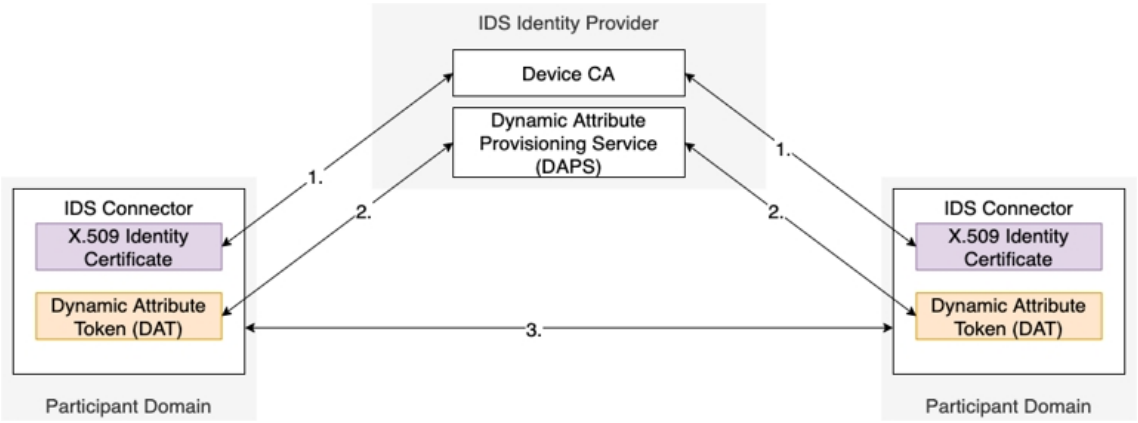
This verified identity is then used to confirm that a participant is a verified participant of the dataspace. This identity providing mechanism is a centralized identity provider, meaning that the identities are provided by one central trusted identity provider. If in the future multiple regional dataspace were to be deployed then it might make sense to switch over to a decentralized identity provider. This would give every participant in the dataspace their own standardized identity which can be transferred and utilized between various dataspaces [52].

When exchanging data the identity interaction between data connectors goes as follows [53]:

1. Each IDS Connector acquires a valid identity certificate from the Device CA (part of the identity provider)
2. Each IDS Connector requests a current Dynamic Attribute Token (DAT) from the Dynamic Attribute Provisioning Service (DAPS)(also part of the identity provider)
3. When establishing communication, the DAT of both IDS Connector instances is exchanged. This is also matched with the used TLS certificate.

For a visual representation of this interaction please refer to figure 2.5.

FIGURE 2.5: Identity Interaction between Data Connectors and Identity Provider in IDS [53]



Data Access Control:
For the purpose of this thesis, access control is the concept of deciding who can access the data a data provider provides within the dataspace. This is crucial for ensuring the data is not access by people who should not be accessing, which is critical when handling personal medical data. In the IDS-RAM this is mostly achieved through the use of usage contracts.

Usage contracts are discussed in-depth in Chapter 8 of 'Designing Data Spaces' [12]. It highlights the importance of usage policies in an IDS as it a crucial building block for data sovereignty and trust in a dataspace. As mentioned above, the policies define how a participants data can be used and by whom. Usage polices are also not static. They can change as the needs of the data provider changes. This adaptability is also important in

the context of a medical dataspace as a patient can change their mind on whether they want to share their data or not.

In order to ensure that the usage policies are followed, there needs to be enforcement of the usage policies in place. In the IDS, this is controlled through monitoring mechanisms to ensure compliance with the enforcement policies. If any such violation is detected, then these offences are actionable. If this is in effective operation it will lead to increased trust within the IDS.

2.4 Data Privacy and Compliance

In this digital age, the importance of data privacy cannot be overstated. Due to this, during the development of the architecture, data privacy and compliance to legal frameworks need to be taken into consideration at all times. This is especially true given the sensitive nature of medical data. The previous section already touched on how usage control is ensured within the IDS reference architecture; meanwhile, this section goes deeper into how data privacy is to be ensured in the context of a regional medical data hub in general and what legal aspects to take into consideration. This section is further divided into three sections, these being Legal and Regulatory Frameworks, Consent Management & Data Anonymization and Risk Management & Incident Response.

2.4.1 Legal and Regulatory Frameworks

Understanding and adhering to legal or regulatory frameworks is crucial to ensure data privacy and compliance within a regional medical dataspace. This section explores both national and European laws and regulations and analyze how they affect the design of the dataspace architecture. These regulations have a foundational role in the final design as the final artifact must follow these regulations to be functional in a practical setting.

European Legislation:

When discussing data privacy within the EU, the **General Data Protection Regulation (GDPR)** is sure to get mentioned quickly. This is logical because the GDPR has had large affects on all organisations which hold data on EU citizens since it was introduced in 2016.

The General Data Protection Regulation (GDPR) is thus a well-known EU regulation [54]. The law aims to protect people through rules regarding the processing and movement of personal data. It protects fundamental rights and freedoms of people, more specifically their "right to the protection of personal data" [55].

The GDPR is a regulation, this is a form of EU legislature which means that it is a binding legislative act which must applied in all EU member states. It is thus stronger than a EU directive, which is a piece of legislature which sets out a goal which EU countries have to achieve with their own legal system [56]. A regulation does however allow countries to add to the given directive. In the Netherlands this is done through the UAVG.

However, the GDPR does not only apply to organisations which are based in the EU. Since the aim is to protect EU citizens the law applies to any company or organisation

which processes data of EU citizens. This means that organisations from outside of the EU, which do business in the EU also have to follow the rules put forth in the GDPR [57].

Along with the creation of the GDPR, a European Data Protection Board (EDPB) was also formed. It is composed of representatives from national data protection authorities (Autoriteit Persoonsgegevens in the Netherlands) and the European Data Protection Supervisor. The tasks of the EDPB is to provide guidance regarding the GDPR and advising the European Commission on issues related to personal data. The fines are issued by the national authorities [58].

Organisations which severely violate (listed in article 83(5)) the GDPR can receive very large fines. The fine can be up to 20€ million euros, or 4% of the annual worldwide turnover, which ever is greater [59]. The largest GDPR fine which have been issued so far was to Meta, they received a 1.2€ billion fine [60].

This background on the GDPR leads to what the architecture design needs to feature, in order to not break the regulations put forth by the GDPR. The rest of this section on the GDPR highlights articles in the GDPR which are relevant for the medical dataspace architecture. All information is extracted directly from the GDPR legislation document [54].

- **Article 5:** Article 5 covers the principles relating to processing of personal data. It contains seven of these principles each pertaining to a different aspect of the processing of personal data. The first principle states that personal data shall be processed lawfully, fairly and in a transparent manner. Lawfulness is defined in article 6 and has to comply with one of 6 conditions. These conditions is described in more detail below. Fair in the context of article 5 can be summarized as a data processor should only be processing and handling personal data in ways that the individual would expect. Last, Transparency means that it is clear to the individual how you collect and manage data.

  The second principle of article 5 is the principle of purpose limitation. Purpose limitation means that data cannot be used for a purpose, different to the purpose for which the individual has given consent. If the data processor intends to use a data for a different purpose, new consent has to be obtained.

  The third principle is that of data minimisation. This entails that the data processor does not store more data than which is required for the purpose for which it is intended.

  Accuracy is the fourth principle listed in article 5. In order to comply with this principle the data processor has to ensure that personal data is accurate, and fit for the purpose for which it is intended.

  The fifth principle is storage limitation. Storage limitation within this context means that data cannot be stored for longer than is required for the purpose of the data, and there needs to be justification for storing the data. The principle does however contain an exception to this rule. Personal data can be held for longer if it is being stored for: public interest archiving, scientific or historical research or statistical

purposes.

The last principle is the principle of integrity and confidentiality. The principle enforces that the data is secured in an appropriate manner against "unauthorised or unlawful processing and against accidental loss, destruction or damage".

- **Article 6:** The processing of data is considered lawful if it follows at least one of the following conditions:

  1. The data subject has given consent to the processing
  2. The data is required for a contract to which the data subject has agreed
  3. The processing of the data is required to comply with a legal obligation of the data processor
  4. The processing is necessary to protect the vital interests of the data subject or another natural person
  5. The processing is necessary for the performance of a task which is carried out in the public interest
  6. The processing of the data is only allowed if the interests of the data processor do not override the fundamental rights of the data subjects.

- **Article 7:** Article 7 is about giving of consent. The article has four paragraphs, each of which are conditions for consent:

  1. The data controller is to be able to demonstrate that the data subject has given consent to processing of their personal data.
  2. If a data subject gives consent to the processing of their data within a large context which also concerns other matters, then the consent for the processing of data is to be presented separately from the other matters.
  3. The data subject has the right to withdraw the data. All the processing done before the withdrawal of data is still considered lawful and withdrawing of consent is to be as easy as giving it.
  4. When deciding if consent is freely given, it's important to check if someone was forced to agree to something unnecessary to get a service or fulfill a contract. If an organisation requires you to share personal data that is not deemed needed for the service, consent might not be considered freely given.

- **Article 9:** Article 9 concerns special categories of personal data, which require extra protection. One of these categories is the category of health data, and given the nature of this thesis, special care must be given to this article. The article states that the processing of this data is prohibited unless certain conditions are met. Ten such conditions exist. The conditions most relevant for the medical dataspace are as follows, the processing of data is not prohibited if:

  - the data subject has given explicit consent to the processing of their medical data.
  - processing is necessary to protect the vital interests of the data subject or of another natural person where the data subject is physically or legally incapable of giving consent. This is relevant for the medical dataspace as patients might not always be in a state in which they are able to provide consent.

- processing medical data is allowed when it is necessary for purposes such as preventive/occupational medicine, medical diagnosis, providing health or social care or treatment, and managing health or social care systems and services. However, this processing can only be done based on (in this thesis) Dutch law or a contract with a healthcare provider.
- processing is necessary for reasons of public interest in the area of public health, such as protecting against serious cross-border threats to health or ensuring high standards of quality and safety of health care and of medicinal products or medical devices.
- processing is necessary for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes.

  As can be seen in the list above, while medical data does hold stronger protection under the GDPR than other personal data which can be shared online, the GDPR also holds various paragraphs which allow for data exchange in the healthcare sector.

- **Article 13 and 14:** Article 13 and 14 share similarities as they both cover transparency and informing data subjects on how their personal data is being processed. Article 13 applies when data is collected directly from the subjects and requires organisations to provide the subject with information on their identity, the purpose of their data collection, recipients and the rights of the data subject. Article 14 covers similar aspects but is instead focused on data gathered not directly from the data subject but through other sources. Article 14 also adds the requirement for organisations to inform the data subject what other source their data was obtained from. The architecture should thus have elements which allow patients to view the information above in order to ensure trust and transparency.

- **Article 15:** Transparency is also the main objective of article 15. Article 15 ensures that data subjects have the right to find out whether their personal data is being processed. If their data is indeed being processed they have the right to receive information on various aspects of this process. Similarly to article 13 and 14 above, the architecture should have a mechanism for data subjects to have access to this information.

- **Article 17:** Article 17 contains the right to be erasure, commonly known as the right to be forgotten. This article allows for personal data to be erased following certain ground of appeals. These include situations where the data is no longer required for the purpose for which it was collected, or the subject withdraws consent for the processing of the data. However, medical records are a more complicated matter as healthcare providers have a obligation to maintain accurate records to ensure appropriate care. Given that the right to erasure is not absolute (it is based on the conditions mentioned above), requests for erasure of medical records is evaluated on a case-by-case basis.

The GDPR is a very comprehensive legal document and covers a lot more content than described above. That being said, the articles highlighted in this section cover the aspects which are to be included in the architecture in order to comply with the GDPR. Furthermore, the GDPR allows member states to expand on it on a national level. These national regulations is covered in a later section. For more information on the GDPR and

all additional articles please refer to the GDPR itself at [54].
Outside the GDPR, there are more EU legislative components which need to be acknowledged. The EU Data Governance Act aims to facilitate the (re)use of public data, both personal and non-personal. It does this through providing a framework and intermediation services. It is not very relevant for the medical dataspace and thus is not worked out in more detail, but given the focus of the Data Governance Act is to facilitate data sharing in the EU, it was mentioned here.
Lastly, the EU introduced the AI Act in 2024. It is the first piece of legislation in the EU on AI. it Is a legal framework which addresses the risks which AI brings. While the dataspace currently does not feature the use of AI, it could expanded into AI in the future, and then, this legislation would become very relevant.

National (Dutch) Legislation:

Evidently, not only European legislation applies to a regional medical dataspace in Twente. Dutch law also applies and there is a range of legislation which needs to be taken into consideration when developing the dataspace architecture. Below the legislature is described and discussed how these apply to the dataspace where applicable.

- **Wegiz:** The "Wet elektronische gegevensuitwisseling in de zorg" is a law which has been introduced in the Netherlands in July 2023. The law enforces healthcare providers to exchange healthcare data. The motivation behind this law was that despite an increasingly more digitized health sector, health professionals would still be performing repetitive tasks re-entering data, and patients often had to explain their situation multiple times. This example of data not being exchanged between healthcare providers motivated the Dutch government to create a law to expedite the development of health data sharing. [61] The law does not define what has to be shared, instead after various healthcare organizations decide on a data sharing standard (norm in Dutch), this standard is then enforced in the law [62]. This law does not directly influence the architecture as the standards are not agreed upon (yet) and thus no change is needed.

- **UAVG**: The "Uitvoeringswet AVG" is the Dutch national implementation of the GDPR (The GDPR is discussed in more detail in the section below on EU regulations). The GDRP leaves somes elements of the regulation free for EU-member states to arrange themselves, that is what the UAVG does [63]. Amongst other things, the UAVG ensures that children under the age of 16 can not consent to sharing their data. Additionally, it gave the Autoriteit Persoonsgegevens more disciplinary power. Because of the former, the architecture needs to ensure that there is mechanism for parents or guardians to allow for the medical data sharing of their child or ward.

- **Wabvpz:** The "Wet aanvullende bepalingen verwerking persoonsgegevens in de zorg" is responsible for ensuring the use of the Burgerservicenummer (BSN) in healthcare. Furthermore, it gives patients the right to digitally view their own medical file, receive an extract from it and be informed of who viewed their file, when they viewed their file and who gave them to access to their file [64]. The law ensures that the architecture allows for the patients to view their own file, and the other related information described above.

- **NEN-normering 7510, 7512, 7513:** The "Nederlands Normalisatie-instituut" (Royal Netherlands Standardization Institute) manages and maintains standards in

the Netherlands. Some of these standards apply to Dutch healthcare organisations, and thus can be relevant for this report. According to the Ministry of Health, Welfare and Sport the three most relevant for healthcare data are:

  - **NEN 7510:** NEN 7510 is focused on information security in the health sector. It is based upon European ISO standards (27001, 27002 and 27799) and describes demands and also measures organisations can take in order to ensure that private health data stays safe [65].
  - **NEN 7512:** NEN 7512 is a further specification of NEN 7510, and thus, also focuses on the domain of information security in healthcare. It deals with communication between healthcare providers, institutions, insurers and patients. It ensures that the various parties have to provide each other with assurances regarding data exchange [66].
  - **NEN 7513:** NEN 7513 focuses on logging. In this context logging refers to the action of logging who views the private information present in a medical file, and what actions were then taken. The goal of the system is to keep the complete collection of events across various systems and domains verifiable [67].

  For the architecture to comply with the NEN-normeringen it needs to have robust security systems. While the IDS-RAM features some security systems, special care needs to be taken to ensure it passes all security demands of NEN7510 and NEN7512. Further more, the logging system of the IDS-RAM needs to feature in the architecture to ensure compliance with NEN7513.

- **WGBO:** De "Wet op de geneeskundige behandelingsovereenkomst" in its core is slightly further removed from healthcare data than the above regulations. However, there are still some relevant aspects to this law. The law includes the fact that the patient has the right to receive information about their situation, in understandable language and what the treatment is (and what the risks are for that treatment). Furthermore, the healthcare professional is required to keep a file on the treatment and the patient, and the patient has the right to view their file [68].

- **Wet BIG:** De "Wet op de beroepen in de individuele gezondheidszorg" is a law which places rules and regulations on people working in healthcare for a wide variety of professions. The aim of the law is to protect patients from harm by the hand of unqualified healthcare personel. It does this by creating education requirements, ensuring that certain medical procedures can only be done by certain level of healthcare worker and ensuring doctor-patient confidentiality. It also includes punishments for people who break (medical) law. Under this law, all people working within the professions which qualify for this law receive a personal identifier, a BIG-nummer [69].

2.4.2 Consent Management & Data Anonymization

Effective consent management is a cornerstone of data privacy, ensuring that individuals maintain control over their personal health information. This section analyzes mechanisms for obtaining, managing, and documenting patient consent. Additionally, data anonymization techniques are crucial for protecting patient identities while enabling data sharing and analysis for secondary research purposes. This sections thus explores various anonymization methods, their effectiveness in preserving privacy, and their potential use for the

medical dataspace.

Consent Management:

In section 2.3 the matter of consent in the IDS-RAM was briefly touched upon. In this section consent management in the IDS-RAM is linked to how consent management in currently sorted in general in the Netherlands. These two aspects are combined in the architecture in order to satisfy both the requirements of the IDS-RAM while being able to be used in practice in the Netherlands.

The IDS-RAM manages consent control through usage contracts. These usage contracts are constructed of several usage policies . These policies describe permissions and obligations for an IDS resource. In other words, they describe who can use a resource and for what. Usage policies are written in a machine-readable format. This guarantees that the usage contracts can always be processed as defined in the process layer. The architecture needs to feature some method for patients to define their usage policies which can then in turn be constructed into machine-readable usage contracts so they can be followed within the dataspace.

In order to build these usage policies the wishes of the patient in terms of medical data sharing need to be recorded. Rather than create a entirely new system for this dataspace architecture it would be more efficient to lean on existing systems in the Netherlands. A discussion on the topic of consent management in the Netherlands with the lead architect at MST in Enschede lead to a couple of aspects which are relevant for the to-be architecture.

For a long time and up until recently, consent for medical data sharing was acquired separately at every single healthcare provider. In practice, this meant that a patient had to answer the same questions, and provide that consent multiple times at different points along their care process. This is inefficient and can lead to frustration for the patient. A recent development is that of Mitz [70]. Mitz is a consent management system in the Netherlands where patients have control over their own medical data sharing consent. They decide which healthcare providers can share their personal data, and which cannot. They can do this through logging in on a portal on the Mitz website and indicating their preferences. Furthermore, a healthcare provider can change consent during an appointment with the patient if the patient allows the healthcare provider to do so. This allows for patients who are not able to do so themselves, to still give consent for sharing their medical data.

Mitz is still in a relative early stage of development, with the first healthcare providers using it in their operations starting in December 2023 [71]. That being said, the large advantage using Mitz for consent management for both patients and healthcare providers makes it a powerful tool to be used in the architecture. Mitz in the architecture is used for creating the usage policies, which in turn can be used to craft usage contract to be deployed in the dataspace.

Data Anonymization:

For the secondary use of data, an important aspect of the data availability is that of the application of data anonymization. Data anonymization is the process where 'data is transformed in such a manner that formal guarantees, e.g. regarding the risk of singling out, linkage or inference, can be provided' [72]. This procedure is very common in healthcare due to the very sensitive nature of the data in question. Because data anonymization is

such a common technique for medical data, many publications have been published over the years. In 2022 a systematic review was published on data anonymization methods and tools in the healthcare sector. This paper [73] provides a comprehensive overview of current anonymization methods and tools specifically for the healthcare sector, and also analyzes challenges and benefits of these techniques. The paper found 32 publications which were relevant for their purpose and in them, 40 different anonymization techniques were identified, either as main focus of the paper, or mentioned as an alternative. Below you will find the most common techniques (main focus of at least three publications in the review) according to the paper.

- k-anonymization
- l-diversity
- t-closeness
- Cryptographic algorithms
- Pseudonymization
- ϵ-differential privacy

2.5 Interoperability in healthcare

Interoperability is potentially the core of a dataspace. The goal of a dataspace is to facilitate the exchange of data between parties in the dataspace. For this to happen interoperability needs to be achieved on both a semantic and technical level. Both these aspects of interoperability is discussed below. Furthermore, the IDS-RAM also focuses on facilitating interoperability. This too is covered below.

2.5.1 Semantic Interoperability

Semantic interoperability is defined as *'a shared vocabulary and semantics of knowledge exchanged during collaboration'* [74] or in other words, assigning same meaning to data by both parties. This is essential within healthcare as wrongful interpretation of information can potentially have dangerous consequences.

Semantic interoperability in healthcare has been seen as a challenge, being studied regularly, for example in [75] and [76]. An EHR with complete semantic interoperability allows for any healthcare provider to interpret an EHR no matter what organisation compiled it [77]. A paper from 2022 by De Mello et. al. [78] provides a systematic literature review on semantic interoperability in EHRs. This review provides valueble insights which can aid in the development of the dataspace architecture in section 4.

The paper by De Mello aims to highlight the current state-of-the-art of semantic interoperability in (electronic) health records. It does this by looking at what health standard and terminologies are being used. Below you will find summarized versions to both of these aspects of semantic interoperability in healthcare.

Health standards adopted in the studies: The review shows that while there is no true consensus when it comes to standards used in healthcare, many organisations deploy a dual model approach. Usually openEHR with either HL7 or ISO13606. In this setup,

OpenEHR is used to structure the EHR data and transfer of data is facilitated by one of the other two.

Health terminologies used in the studies: The goal of the terminology is to create a common language used by healthcare professionals. To complete this goal, various terminologies have been developed over the years. The paper shows that the most cited terminology is SNOMED CT. Aside from being the most cited, the paper also mentions that SNOMED CT is clinically validated, semantically rich and a controlled vocabulary. ICD is also often quoted and discussed, it focuses on classification of diseases.

Challenges: The paper also discusses challenges present in the field of medical semantic interoperability. The most common challenge mentioned is the problem of unstructured or semi-structured data. While having a open text field is easy to use for a healthcare professional, this is not easily readable for a computer. Legacy systems also shown to create a challenge, as they are not compatible with exchanging data at times. Another challenge inherent to the field of medical semantic interoperability is the challenge of aligning various vocabularies to a common meaning. While the most common vocabularies are mapped to each other, not all existing standards have the feature.

Conclusion: The paper by De Mello et. al. [78] provided valuable insights which can be applied to the dataspace architecture. Using HL7 in the dataspace makes sense as it is common in the field and is already deployed within the central hospital in the Twente region where the dataspace would be operational. Furthermore, SNOMED CT is to be used due to the prevalence in the field and the aforementioned benefits of that terminology system. Also, the current terminology used by the nursing staff at the hospital is mapped to SNOMED CT.

2.5.2 Technical Interoperability

The next level of interoperability to discuss is technical interoperability. Technical interoperability, is the ability for two systems to interact and exchange data on a technical level. Or as defined by ETSI in [79], *'Technical Interoperability is usually associated with hardware/software components, systems and platforms that enable machine-to-machine communication to take place. This kind of interoperability is often centred on (communication) protocols and the infrastructure needed for those protocols to operate'*

Communication protocols: As discussed before in section 2.2.1 **HL7 FHIR** is a communication standard which is very common in healthcare, including in the Twente region.

2.5.3 Interoperability in the IDS-RAM

The IDS Reference Architecture Model contains several features aimed at facilitating interoperability between systems, on both a semantic and technical level.

First of all, the reference architecture being published openly, in itself is a step towards ensuring interoperability. All partners have access to the same documentation, and the IDSA publishes various repositories on GitHub [80]. This open manner of sharing information allows all participating parties of the dataspace to develop their system from the same foundation. This aids in building interoperability from the start.

Next, the IDS-RAM places a lot of emphasis on ensuring that data is described using metadata (more on that topic in section 2.6). Well-described data allows partners in the dataspace to more easily identify the data they are looking for, increasing understanding of the data being shared and thus, interoperability.

Third, vocabularies are a cornerstone within an IDS dataspace. They are a key tool in ensuring semantic interoperability. A vocabulary is decided upon for a dataspace. This vocabulary is then used to define the semantics of the data assets in the dataspace. In order to ensure that all participants in the dataspace have access to the vocabulary, the vocabulary is to be stored in a vocabulary hub, which is accessible by all participants using a browser-based UI and API. According to the IDS-RAM, these vocabularies are also able to be edited by participants. However, this does come with the requirement of appropriate user management to avoid abuse. As discussed above, the standard vocabulary for the regional, medical dataspace is SNOMED CT.

Lastly, the IDS-RAM features the concept of "Standardized Interoperability". Standardized interoperability revolves around the IDS Connector component. The interaction central to this is already discussed in section 2.3, but some details not mentioned there are still relevant here. The connectors in the IDS-RAM are run on participants own devices. Furthermore, Users of connectors need to be identifiable, manageable and all interactions are to be logged. Incidents surrounding the connectors should be reported automatically. Connectors are also to be uniquely identifiable.

2.6 FAIR Data principles in the IDS-RAM

This section on the FAIR data principles in the IDS-RAM starts with a more detailed explanation of what the FAIR data principles are, as the earlier introduction in section 1.1 was rather brief. The FAIR data principles were introduced in 2016 [13] to '*improve the infrastructure supporting the reuse of scholarly data*'. The FAIR principles emphasise the ability for data to be machine readable. The principles are:

- **F**indable: Principles aimed at making available data easier to find.
    1. (meta)data are assigned a globally unique and persistent identifier;
    2. data are described with rich metadata;
    3. metadata clearly and explicitly include the identifier of the data it describes;
    4. (meta)data are registered or indexed in a searchable resource;
- **A**ccessible: Principles ensuring that when data are discovered, it is available to the user (where allowed).
    1. (meta)data are retrievable by their identifier using a standardized communications protocol;
        1.1. the protocol is open, free, and universally implementable;
        1.2. the protocol allows for an authentication and authorization procedure, where necessary;
    2. metadata are accessible, even when the data are no longer available;
- **I**nteroperable: Principles focused on making data more interoperable between systems

1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
2. (meta)data use vocabularies that follow FAIR principles;
3. (meta)data include qualified references to other (meta)data;

- **R**eusable: Principles ensuring that data is made available in a way that allows it to be easily reused by others for future usage.

1. (meta)data are richly described with a plurality of accurate and relevant attributes;
   1.1. (meta)data are released with a clear and accessible data usage license;
   1.2. (meta)data are associated with detailed provenance;
   1.3. (meta)data meet domain-relevant community standards;

For more information on the interpretation and implementation considerations of the principles above, please refer to [81].

Now that the principles are discussed, this thesis now analyzes how the IDS-RAM applies these principles in the reference architecture. This is done for each of the four principles described above, where possible, linking a feature of the IDS-RAM to a specific (sub)principle.

- **Findability:**
  - Various elements (eg. Connector, data asset or participant) within the dataspace are identified using a unique identifier. (F1)
  - Data providers are to provide a (meta)data broker with metadata about the data they are providing, the broker in turn ensures that the data can be found by other participants. (F2)
  - Connectors provide self-descriptions which contain information about the connector like organisation responsible for maintaining the connector and content and type of data provided via the connector. This self-description can also include the identifiers of that data. (F3)
  - (Meta)data is registered and brokered via the IDS broker in the dataspace. The broker then presents a searchable resource. (F4)
- **Accessibility:**
  - IDS Connectors use standardized protocols for data access, ensuring consistent retrieval. (A1)
  - The IDS publishes their protocols openly, and thus everyone is free to implement them. (A1.1)
  - The RAM features various avenues for ensuring authentication, authorization and access control. (A1.2)
  - While the IDS-RAM seemingly does not necessarily state anything regarding maintaining metadata after deletion of the data. There are aspects of the IDS-RAM related to it. It features record keeping of all data transactions which can be kept available after deletion of the data. Furthermore, the broker can be configured in a way where the metadata is retained. (A2)

- **Interoperability:**
  - The IDS features a readily available shared knowledge representation in the form of the IDS Information layer. The role of the information layer is to specify a common vocabulary for the dataspace. (I1)
- **Reusability:**
  - The data which is provided by data providers is described by (meta)data extensively before being provided to the broker. This can include the relevant attributes described in the FAIR data principles (R1)
  - The data usage agreements are a core element of the IDS-RAM using the usage policies and contracts described in section 2.3. (R1.1)
  - Provenance of data is a topic which is again a core element of the IDS-RAM. This is mainly done through the Clearing House. The Clearing House is a node within the dataspace responsible for logging all data transactions within the dataspace. These logs can then be used to provide data provenance. (R1.2)

To conclude, the above shows that the IDS-RAM has good coverage of the FAIR data principles. While not all principles are directly discussed within the IDS-RAM, the architecture can certainly be used as a foundation for a dataspace which follows the principles of FAIR data management.

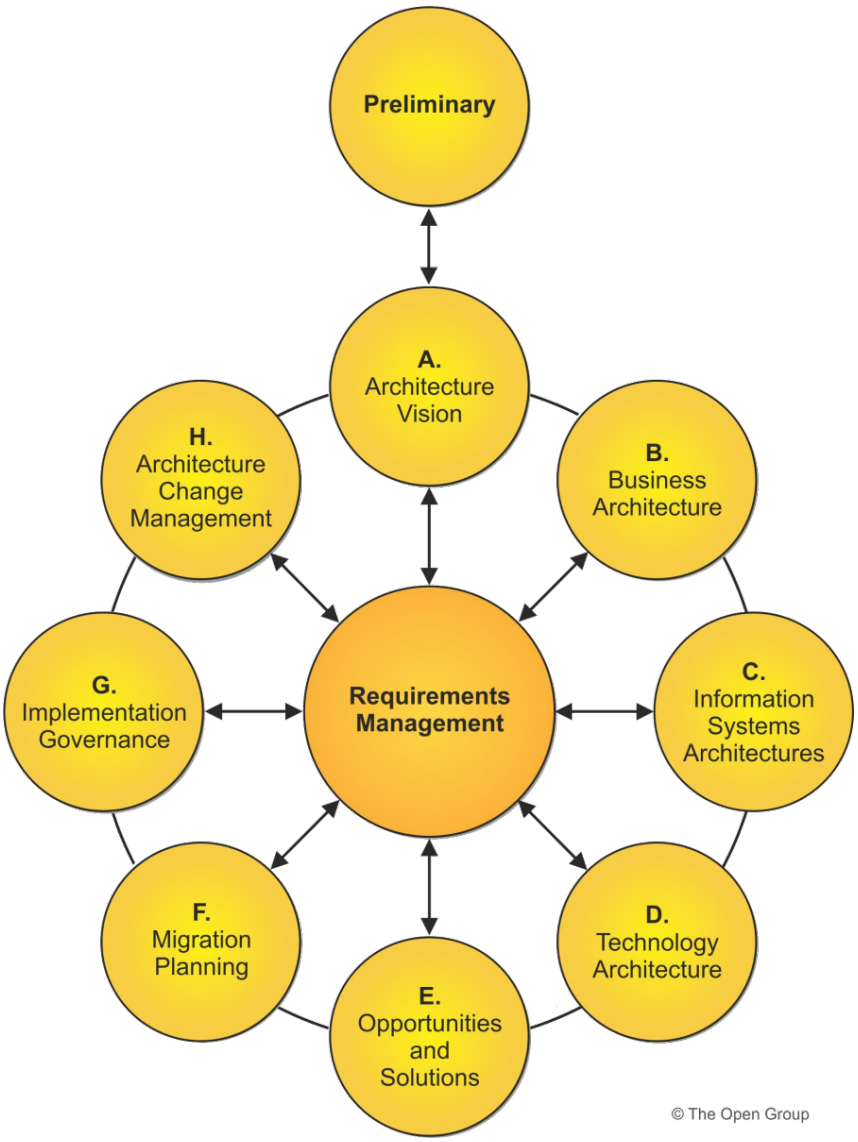
2.7 TOGAF

The Open Group Architecture Framework (TOGAF) is a comprehensive and very widely adopted [82] methodology for developing enterprise architecture models. It provides a structured approach for the entire process from the preliminary stages to the governance of the architecture following deployment.

Central to the TOGAF methodology is the Architecture Development Method, or ADM for short. This ADM features eight stages (see figure 2.6 by the Open Group) and provides guidance for the process of developing and managing the enterprise architecture lifecycle [83]. The steps in the ADM are:

1. Architecture Vision
2. Business Architecture
3. Information Systems Architectures
4. Technology Architecture
5. Opportunities and Solutions
6. Migration Planning
7. Implementation Governance
8. Architecture Change Management

FIGURE 2.6: The stages of the ADM



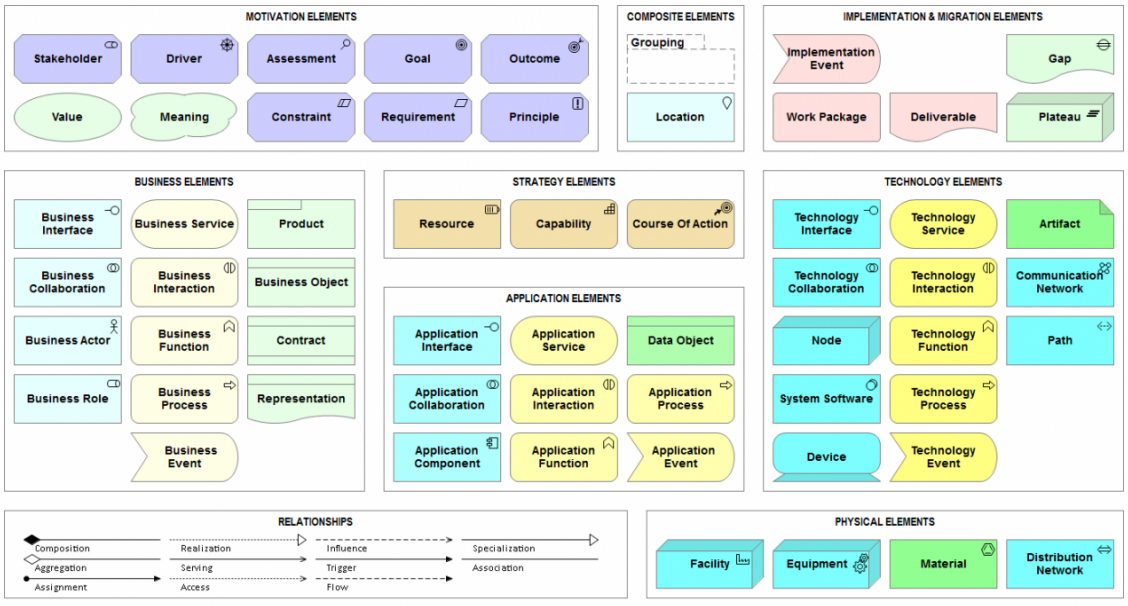
Along with these eight steps there is a preliminary stage which triggers the cycle and it is continuously supervised by Requirements Management, which is visualized in figure 2.6 by bidirectional arrows between requirements management and all the other stages.

The ADM is also seen as modular, allowing the architect to choose which stages are most relevant to the project they are working on. It is also designed to be an iterative process, where revisions can be made to earlier stages later on in the project. Due to this thesis focusing on the development of the architectural model of the regional, medical dataspace, the first four stages is fully developed, analyzed and discussed.

2.7.1 ArchiMate

For this thesis the use of the ArchiMate modelling language [14] was selected. ArchiMate is a flexible language and has numerous advantages, for example, ArchiMate allows for a comprehensive view of the entire enterprise (or dataspace in this thesis), it goes beyond just the application layer and it is used by many organisations [84]. Furthermore, given that both TOGAF and ArchiMate are developed by the same organisation (The Open Group), they are inherently designed to complement each other, ensuring seamless integration and consistency between the architectural framework (TOGAF) and the modeling language (ArchiMate). For an overview of what the ArchiMate language looks like, please refer to figure 2.7.

FIGURE 2.7: Overview of the ArchiMate language



Chapter 3

Stakeholders, Goals and Requirements of the Dataspace

After all the background information provided in chapter 2, this chapter analyses and discusses the stakeholders, goals and requirements of the regional, medical dataspace. After this, a conclusion for chapter 2 and chapter 3 includes the identified problems of the subject matter, and the answers to research question 1.1, 1.2 and 1.3.

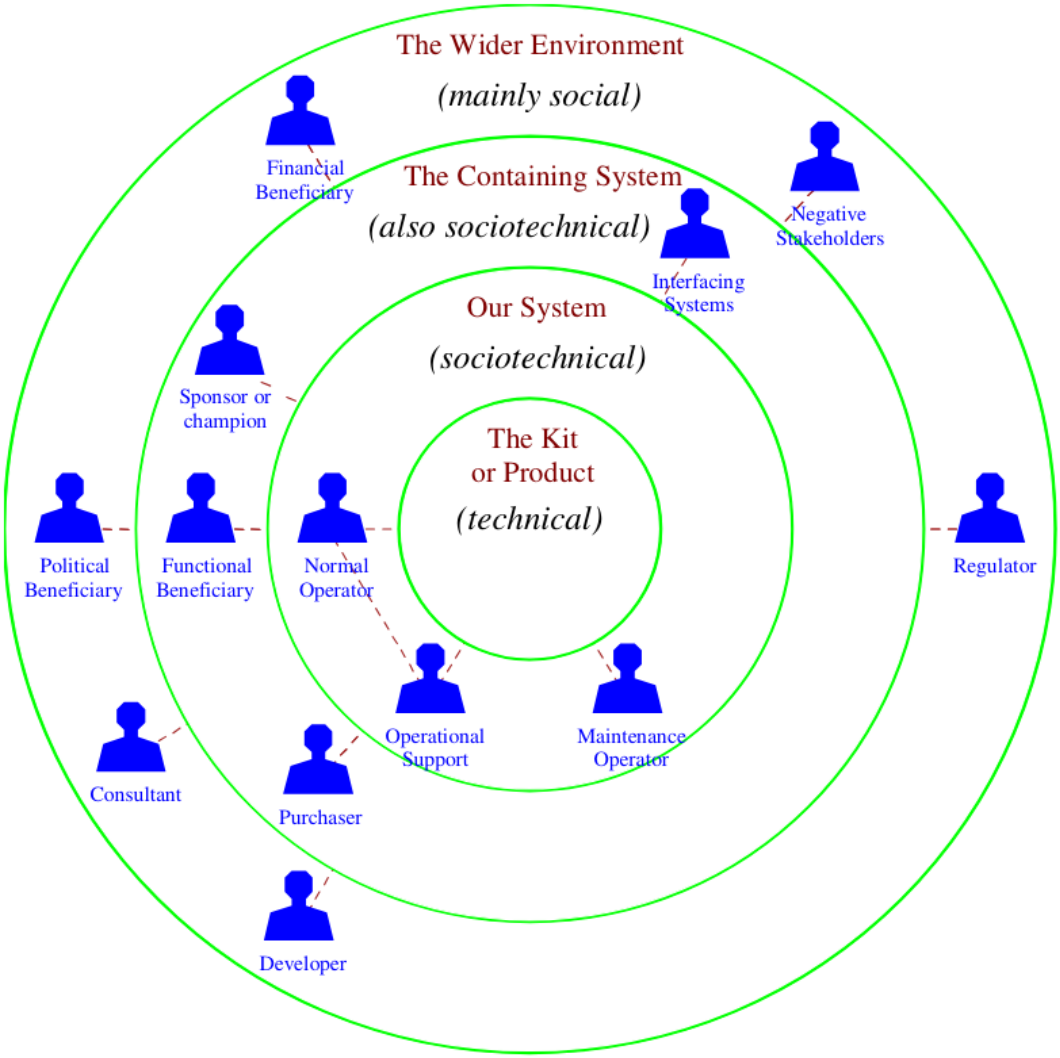
3.1 Stakeholders within the Regional Health ecosystem in Twente

Wieringa defines a stakeholder as: '*a person, group of persons, or institution affected by treating the problem*'. It is relevant to identify the stakeholders for this project in order to determine what their goals are for the project, and how their situation would change if the project proves successful. After the identification of the stakeholders they also receive a role according to the taxonomy described by I. Alexander in [85].

Alexander also describes a conceptual framework which can be used to identify stakeholders. This framework is based on an onion model and is very useful in identifying how close a certain stakeholder is to the product. This onion framework can be found in figure 3.1. Following the identification of the stakeholders of the dataspace, these are then mapped onto this onion model for a comprehensive overlook of the stakeholders involved in the project. For more information on the details of the conceptual framework, please refer to chapter 3 of [85].

In the table (table 3.1) below you will find the identified stakeholders, ordered by the layers of the onion model, with their role according to the taxonomy of Alexander.

FIGURE 3.1: Onion Diagram of Product Stakeholders from [85]



Stakeholder	Stakeholder role	Stakeholder role description
Healthcare professionals	Normal Operator	Give routine commands to the artifact, sometimes called "end users".
Researchers	Normal Operators	Give routine commands to the artifact, sometimes called "end users".
ZorgNetOost IT staff	Maintenance Operators	Support normal operators in their use of the system and help to keep the system operational.
Health Insurance Companies	Sponsor	Initiates and provides a budget for developing the artifact.
Patients	Functional Beneficiary	Benefit from the output produced by the system, sometimes called "users" of the artifact.
Healthcare organisations	Functional beneficiary	Benefit from the output produced by the system, sometimes called "users" of the artifact.
ZorgNetOost	Functional Beneficiary	Responsible for ensuring the system is operating, and further development of the artifact
IDSA	Consultant	Support development of the artifact.
Developers	Developers	Such as requirements engineers, designers, programmers, and testers build the system. They are not normal operators of the system and do not benefit from its output during normal operation.

TABLE 3.1: Stakeholders and their roles for the dataspace architecture

Table 3.1 features stakeholders specific to the Twente region, like ZorgNetOost. While this is done because the use case of this thesis focuses on the Twente region, one could potentially apply the identified stakeholders to other regions as well. However, caution must be taken, as the role ZorgNetOost fulfills in Twente will also need to be addressed in other regions.

3.2 Goals

Following the identification of the stakeholders it is also relevant to determine their goals behind the project, in order to conclude what motivates them to be involved in the project. Direct involvement also varies between each stakeholder, so level of involvement is also indicated. You can find this information in table 3.2. The goals can be grouped into 5 categories these are:

1. Data availability
2. Process efficiency
3. Secondary use of data
4. Regulatory compliance
5. Goal not directly related to architecture, but some goal for the development

Stakeholder	Involvement during development	Goal	Goal Category
Health Insurance Companies	Provides funding for the project, and when implemented would be one of the largest data providers and users	To make data more readily available and make their processes more efficient	1, 2
Patients	No involvement in development	Their data would be more available to care providers, resulting in a more efficient flow through the healthcare system	1, 2
Researchers	No direct involvement in development. Could be potentially asked to provide requirements for the secondary use of medical data	To have more data available to them for research	3
Healthcare organisations	No direct involvement in the project	Would be both data provider and data user. Similar to MST they would hope to improve their processes	2
IDSA	Provides the reference architecture	No goal for this project	5
Healthcare professionals	No involvement in development	Their work process will improve	2
Developers	High involvement in the development of the project as they develop the project following the architecture	Their goal is to deliver a successful final product	5
ZorgNetOost	High involvement in the development of the project as they are responsible for the development	To ensure that a data-sharing initiative is in place which fulfills the demands of the Integraal Zorg Akkoord	4
ZorgNetOost IT staff	Little direct involvement in the project	No goal in the project, but it would become part of their work responsibilities and thus, would desire it to be a stable system	5

TABLE 3.2: Stakeholders with their level of involvement in the project, goal, and goal category

3.3 Requirements

A crucial part of the design cycle is to determine the requirements for the artifact. The requirements quantify the goals of the stakeholders reported in chapter 3. The requirements also provide guidelines when designing the architecture.

Wieringa defines a requirement as '*a property of the treatment desired by some stakeholder, who has committed resources (time and/or money) to realize the property.*' [15](p. 51). This definition ensures that in order to specify accurate requirements, stakeholders have to be consulted. Furthermore, literature on dataspaces and the IDS-RAM is analyzed in order to identify all relevant requirements. Given these various sources of the requirements, there is bound to be some overlap between them. Given this, at the end of section 3.3, there is a cleansed, final list of requirements which are to be covered by the architecture design.

3.3.1 Requirements from existing literature on dataspaces

First, published literature is consulted in order to determine whether literature exists which describe requirements for dataspaces in general. In order to discover papers on these requirements the following query was utilized:

("Dataspace" OR "Data space" OR "dataspaces" OR "data spaces") AND requirements

This query ensures that all common spellings of the concept of dataspaces are covered and specifically looks at papers discussing requirements. When filtering for these terms in the title, abstract and keywords, and for papers published in 2017 or later, resulted in 188 results. When then removing non-English papers, 174 were left.

These papers were then briefly analysed for their relevance. Relevance here is based on whether requirements were a key topic of the report in the abstract. This left 19 papers which could hold requirements for dataspaces. Of these 19 papers 9 mentioned requirements for dataspaces specifically which were relevant for this thesis. In this context that means that these papers discussed requirements which are relevant for dataspaces in general, or dataspaces in the healthcare sector. The nine papers from which the requirements were extracted can be found in table 3.3, the requirements can be found below:

1. **There needs to be an ontology in place for semantic interoperability:** Within the dataspace there needs to be a widely accepted ontology to ensure semantic interoperability between participants.

2. **The negotiation of data exchange needs to be automated:** No human intervention should be required to facilitate the data exchange.

3. **Usage contract creation needs to be facilitated through a graphical user interface:** Given that the system is to be used by a very wide variety of people, a easy-to-use graphical interface should exist for developing usage policies.

4. **There needs to be the ability to transform data to ensure interoperability:** Not all healthcare providers use the same system in their operations. Due to this, data transforming needs to be facilitated to ensure that all participants can work effectively in the dataspace.

5. **Usage policies need to be enforced:** Usage policies only work when they are enforced. If participants break usage policies, there need to be consequences.

6. **Data needs to be described using metadata:** In order to facilitate discovery of data (especially for the secondary use of data) it needs to be described using metadata. This also aids in the data following the FAIR data principles.

7. **In case of malfunction of the dataspace, the relevant stakeholder must be notified automatically:** Given the critical importance of the system when implemented in ensuring that medical data is available for healthcare providers; if something happens the correct stakeholders needs to be alerted in order to rectify the problem.

8. **Upon completion of the dataspace, a legal and organizational compliance review needs to be completed:** Given the personal nature of medical data (and data in general) a formal review of legal compliance must be completed to ensure compliance with all relevant legislation.

9. **An overview of participants of the dataspace must be available:** All participating healthcare providers need to be known.

10. **A user needs to be able to access all relevant data and services, using a single set of credentials**: In order to aid in ease-of-use of the system, all services should be protected by a single set of credentials.

11. **A ticketing system is to be provided for users to report technical issues or get support:** In a large system, problems are bound to happen at some point. In order to rectify any problems which may come up, users need to be able to report problems and these problems are then to be turned into tickets to fix the issue.

TABLE 3.3: Publications from which the requirements were extracted

Publication Title	**Year of Publication**
Approaching a Common Conscious Dataspace from a Data Provider Perspective – Requirements and Perspectives [86]	2022
Policy Patterns for Usage Control in Data Spaces [87]	2023
Towards a Reference Enterprise Architecture to enforce Digital Sovereignty in International Data Spaces [88]	2022
Compliance by Design Methodologies in the Legal Governance Schemes of European Data Spaces [89]	2024
Cybersecurity policy framework requirements for the establishment of highly interoperable and interconnected health data spaces [90]	2024
A Data Connector Store for International Data Spaces [91]	2022
Designing a Reference Architecture for Collaborative Condition Monitoring Data Spaces: Design Requirements and Views [92]	2024
Requirements and Building Blocks for Manufacturing Dataspaces [93]	2023

3.3.2 Requirements from IDS-RAM

Since the architecture is to follow the IDS reference architecture model requirement may be derived from the IDS-RAM or other material published by the IDSA.

One of the documents published by the IDSA is the IDSA Rulebook [94]. This rulebook, as the name implies, aims to describe which rules are mandatory for an operating data space. Part of this rulebook is a description on functional requirements for a data space [95]. This document provides a lot of information, without defining true requirements. The requirements extracted below are generalized requirements from the document to avoid creating an excessive amount of requirements.

12. **Participants in the dataspace have retain data sovereignty:** Data sovereignty is a corner stone of the IDS-RAM. This means that the data owner has full control over how their data is used, shared, and managed within a dataspace.

13. **The dataspace needs to have defined policies which specify what attributes an applicant must have to become a trusted participant in the dataspace:** There need to be defined policies with healthcare providers have to obtain to become a trusted partner in the dataspace.

14. **Data discovery through metadata needs to be available to participants in the dataspace:** In the IDS-RAM, discovery through metadata is a crucial aspect of data discovery in general. Each IDS-RAM following dataspace as such, should have the facilities for data discovery through metadata.

15. **A common vocabulary amongst participants needs to be agreed upon in order to ensure semantic interoperability:** See requirement 1.

Next, with the IDS-RAM being a reference architecture it also provides ample requirements. These requirements can be summed up by formulating a single requirement (being that the to-be architecture has to follow the IDS-RAM). However, this does not provide the specificity required for designing an architecture. The requirements extracted from the IDS-RAM can be found below:

16. **The architecture must ensure that the relevant stakeholders have access to the data they require, when they require it:** In this case healthcare professionals need to be able to access the data when they require it.

17. **There need to be security (authentication and authorization) measures in place for participants:** Security measures need to be in place to ensure the data does not end up in unauthorized possession.

18. **Data usage policies need to be created and enforced:** See requirement 5.

19. **Interoperability between standards need to be ensured:** See requirement 4.

20. **Processes need to be described on how to onboard participants, offer data, perform data contract negotiation, exchange data and publish using data apps:** As with any system, adequate documentation needs to exist for any participant to be able to use the system correctly.

3.3.3 Requirements from stakeholders

The requirements described in the two sections above all relate to either dataspaces in general, or the IDS reference architecture. In this section, the requirements are gathered from multiple stakeholders from Twente. These three stakeholders were selected as they all cover a different aspect of the data space. The first is the lead enterprise architect at MST, representing the main hospital in the region and their technical requirements, the second is the CNIO at MST (Chief Nurse Information officer) which is an important primary care stakeholder, the last is a researcher to represent the secondary use of data.

Architect at MST:

Normally, according to Wieringa, a stakeholder is not to define requirements as that is a task for the design researcher. An exception is made however for this first stakeholder questioned about the requirements of the project. This stakeholder works as Lead Architect at MST and thus, has experience with formulating effective requirements. The interview was an open discussion while adhering to certain aspects which were to be covered (Primary workflow, Data Usage, Existing Requirements, Operational). The requirements which were formulated following the interview are:

21. **The healthcare providers are to be able to use their current software:** It is not desirable for healthcare professionals to have to switch, and learn, new software system. The existing systems need to stay available for daily use.
22. **The healthcare provider needs to be able to have access to their patients information when required:** See requirement 16.
23. **The patient needs to be able to determine themselves what parties can share and receive data:** This is related to the creation of usage policies. Patients need to be able to decide on who can access their data, which in turn need to be turned into IDS usage policies.
24. **Usage policies need to be able to be changed in real time:** A patient can be at a doctor and decide to change their permissions in order to allow the doctor they are seeing to access their data. To facilitate this, usage policies need to be able to update in real time.
25. **Updates to patient data can not take longer than 5 minutes to be available to healthcare workers:** Sometimes having the most up-to-date information on a patient can have big consequences for their care. Due to this, updates to patient information cannot take longer than 5 minutes to be updated and visible in the system.
26. **HL7 FHIR needs to be used:** Due to current usage, and future developments, the use of HL7 FHIR is deemed a requirement.

CNIO:

The next stakeholder works as Chief Nursing Information officer at MST. This means that she works both in the primary care role as nurse at the Acute Admissions Unit, but is also responsible for the digitization process of the nursing department within the hospital. Given this role division, the interview focused mainly on the role of data in primary care. The requirements which were collected during this interview are as follows:

27. **Data collected at first entry into the hospital needs to be available throughout the process:** It is not an uncommon occurrence for information which is collected from a patient upon entry into the hospital to be lost. This needs to be rectified in the new system.

28. **Updates to patient data can not take longer than 5 minutes to be available to healthcare workers:** See requirement 25.

29. **Data from the lab needs to also be available to primary care providers:** Results from patient lab work need to also be available to primary care providers.

30. **The system needs to be able to be accessed through both phone (at MST through HiX mobile) and PC:** Primary care workers access EHRs through both phone and mobile PC workstation. These options need to both stay available in the dataspace.

31. **Access to a patient's data is to be dependent on whether a care provider has the rights to that data:** Not all healthcare professionals have the right to access every patients' data. The system needs to have a credential check to ensure only authorized people have access.

32. **Nanda NOC NIC or SNOMED CT is to be used:** Currently the primary care terminology used in the hospital is Nanda NOC NIC. However, efforts are actively ongoing to map Nanda NOC NIC to SNOMED CT. As such, the new system needs to either use Nanda NOC NIC or SNOMED CT.

Researcher:

The next stakeholder to be interviewed is that of a medical researcher in Twente. She works at the Wetenschapsbureau (Research office) at MST. She works as secretary at the office, but also is responsible for research monitoring and quality. Given that easier access to medical data is an important aspect of the medical dataspace, gathering requirements for this secondary use of data is of the utmost importance. Below you will find the requirements gathered from this interview.

33. **Retrospective data needs to stay available for medical research:** Some research projects need data over time. To facilitate this retrospective data needs to be retained (where authorized by the patient).

34. **Research projects must have completed the required paperwork prior to access to the data, and there needs to be a point where this is checked:** Research projects require paperwork to be allowed to do their activities. A check needs to be in place to ensure all required paperwork is present.

35. **Certain research projects require data to be anonymized, there needs to be some way to facilitate this:** Anonymized data allows for research projects to gain the insights they require, without comprimizing privacy. There needs to be a way to access anonymized data.

36. **Patients reserve the right to not have their data used for scientific research, and the system thus needs to facilitate not having those patients' data available for research:** This relates to usage policies. In the creation of usage policies some questions need to relate to medical research using their data.

3.3.4 Summary of requirements

This table below is the final list of requirements as extracted from all the sections above. Some cleaning of the requirements is done to ensure they all follow the same structure and there are no duplicate requirements for the architecture. This final list can be found in table 3.4. Due to the architecture not being implemented in this paper, sadly no non-functional requirements are included. Future research where the implementation is included would allow for non-functional requirements to be described and tested.

Table 3.4: Requirements for the Dataspace architecture artifact

ID	Requirement	Origin
1	There needs to be an ontology and common vocabulary in place in order to ensure semantic interoperability	Literature and IDS-RAM
2	Negotiation of data exchange needs to be automated	Literature
3	Usage contract creation needs to be facilitated through a graphical user interface	Literature
4	There needs to be the ability to transform data to ensure interoperability, and in doing so, maintain interoperability between standards	Literature and IDS-RAM
5	Usage policies need to be enforced	Literature and IDS-RAM
6	Data needs to be described using metadata, allowing for discovery through metadata in the dataspace	Literature and IDS
7	In case of malfunction of the dataspace, the relevant stakeholder must be notified automatically	Literature
8	Upon completion of the dataspace, a legal and organizational compliance review needs to be completed to ensure compliance with all EU and national legislation	Literature
9	An overview of participants of the dataspace must be available	Literature
10	A user needs to be able to access all relevant data and services, using a single set of credentials	Literature
11	A ticketing system is to be provided for users to report technical issues or get support	Literature
12	Participants in the dataspace retain data sovereignty	IDS-RAM
13	The dataspace needs to have defined policies which specify what attributes an applicant must have to become a trusted participant in the dataspace	IDS-RAM
14	The architecture must ensure that the healthcare provider needs to be able to have access to their patients' information when required	IDS-RAM and Stakeholders
15	There need to be security (authentication and authorization) measures in place for participants	IDS-RAM
16	Processes need to be described on how to onboard participants, offer data, perform data contract negotiation, exchange data and publish using data apps	IDS-RAM
17	Healthcare providers need to be able to use their current software	Stakeholders

ID	Requirement	Origin
18	Patients need to be able to determine themselves what parties can share and receive data	Stakeholders
19	Usage policies need to be able to be changed in real time	Stakeholders
20	Updates to patient data can not take longer than 5 minutes to be available to healthcare workers	Stakeholders
21	HL7 FHIR needs to be used	Stakeholders
22	Data collected upon first entry into the hospital needs to be available throughout the process	Stakeholders
23	Lab data must be available to primary care providers (where authorized)	Stakeholders
24	The system needs to be accessible by both phone and PC	Stakeholders
25	Nanda NOC NIC or SNOMED CT is to be used	Stakeholders
26	Retrospective data for research needs to be available (where authorized)	Stakeholders
27	There needs to be a check in place to ensure researchers have completed all required paperwork before being allowed to access data in the system	Stakeholders
28	There needs to be an anonymization process available	Stakeholders
29	Patients need to be able to opt out of their data being used for research purposes	Stakeholders

Over the course of these chapters several aspects of the problem space have been identified and analysed. First the report discussed the current state of medical sharing in the Netherlands by looking at both common medical standards in the Netherlands and current medical data sharing initiatives which are active in the Netherlands. This was then followed by a closer look into dataspaces, and by extension the International Dataspace Association. The IDSA-radar was then used to identify currently operating dataspaces in the healthcare domain. Lastly, stakeholders were identified and their corresponding goals for the project were analysed.

3.4 Conclusion Ch. 2 and Ch. 3

In section 1.4 of this thesis three questions were described to be answered during the problem investigation phase of the design research cycle. These questions are now answered definitively here.

1.1 *What is the current state of medical data sharing in the Netherlands?*
Healthcare providers in the Netherlands generally use standards used worldwide for medical data exchange. The most common standards used in the Netherlands, as in large parts of the rest of the world are HL7 (FHIR) to facilitate data transfer between applications, DICOM for exchanging medical imagery, SNOMED CT for semantic interoperability and OpenEHR for mainting electronic health records. Furthermore, there are more local standards and systems in use like the AORTA/LSP interaction for data exchange between the GP of the patient and the hospital specialist. Lastly, various data exchange initiatives are active within the Netherlands which either cover slightly different aspects of the medical IT field or cover different geographical regions

within the Netherlands. There is currently no nationwide datahub for medical data sharing in the Netherlands however.

1.2 *Which medical dataspaces are currently in operation following the IDSA-RAM?*
While some medical dataspace exist which follow the IDS reference architecture, none of these fulfill the role of the dataspace described in this thesis. The existing dataspaces have a focus which does not align with the regional medical dataspace described in this thesis. They mostly on secondary data or a vastly larger scope in terms of geographical area. They do not aim to create an environment where patients have control over their own data while allowing easy access through the healthcare process for all (authorized) healthcare providers.

1.3 *Who are the stakeholders for the project and what are their goals?*
Stakeholders were identified in section 3.1 and can be found in table 3.1.

Additionally, to complete this section, the problems identified need to be conclusively drawn from the discussed material and listed. These problems are to be verified in section 5.

- **Lack of regional medical data exchange:** Currently, there is no regional initiative for data sharing in Twente. This makes it difficult for healthcare providers to access and exchange data. The current situation mostly allows data exchange between general practitioners and hospitals, while other healthcare providers usually have no data exchange. This leads to inefficiency, as data must be re-collected or manually requested from other organizations.
- **Need for interoperability and data sovereignty:** There is a need for a system that ensures interoperability between existing standards and systems while maintaining data sovereignty as medical data is considered very personal. This is crucial to ensure that healthcare providers have access to data in a usable format and that data owners maintain control over their data.
- **Patient privacy and regulatory compliance:** When designing a system for sharing medical data, patient privacy and compliance with legal frameworks, such as the GDPR and WGBO, must be prioritized. This includes mechanisms for obtaining consent, managing consent, and ensuring secure and confidential data exchange.

Chapter 4

Twente Dataspace Reference Architecture

This chapter covers the design of the dataspace architecture. The architecture is presented following the TOGAF framework. In this chapter research question 2.1 and 2.2 is covered.

4.1 Architecture

In this section all the above preparatory treatment design research is applied in order to develop an architecture of a regional, medical dataspace which satisfies the requirements of the IDS-RAM.

4.1.1 Preliminary Stage

The preliminary phase of the TOGAF ADM is a preparatory phase. It is about defining the "where, what, why, who and how" of the architecture [96]. Given that the previous sections of this thesis already cover the relevant aspects, this preparation is not discussed in further detail.

4.1.2 Phase A - Architecture Vision

Phase A of the TOGAF ADM is called Architecture Vision. This phase includes the defining of the scope and identifying stakeholders, amongst other things. The TOGAF documentation [96] provides steps to complete phase A:

1. **Identify Stakeholders, Concerns, and Business Requirements**
   As the name of this step implies, this step of phase A has the intention to identify stakeholders, along with their concerns/objectives for the architecture. To do this a stakeholder map is to be produced. This stakeholder map has the stakeholder and their objective for the project. This very similar to table 3.1 and thus, is not repeated here. However, an additional task part of this second step is the identification which viewpoints are required to inform all the stakeholders which have been identified.

   - Objective-Motivation Viewpoint: ZorgNetOost, Health Insurance Companies
   - Business Process Viewpoints: Healthcare Organisations, Patients, Researchers
   - Application Viewpoint: Developers, Healthcare Organisations
   - Technology Viewpoint: Developers, ZorgNetOost
   - Layered Viewpoint: All

2. **Confirm and Elaborate Business Goals, Business Drivers, and Constraints**
For this step, the objective is to identify the business goals and strategic drivers of the organization. In September 2022 various organisation in the Netherlands which all together represent most healthcare provided in the country signed an agreement. This agreement, called the Integraal Zorgakkoord (IZA)(Integral Care Agreement) [97], featured many elements for improving healthcare in the Netherlands. Relevant for this thesis however is the agreement that there is to-be a national network for the exchange of medical data (Appendix I of the IZA). Additionally, by 2025, all citizens need to have access to their medical data through an online environment. Given that various healthcare providers in the Twente region also signed[1] this agreement, there is now a concrete business goal for the organisations to improve the state of their data exchange. Constraints are largely related to legal frameworks as discussed in section 2.4.1. Aside from progress being driven by the IZA, there is also always pressure to improve for healthcare organisation in the Netherlands as the industry is suffering from a very high workload for the workforce [98]. Due to this, any improvements on the efficiency of the care processes can help with that problem.

3. **Evaluate Business Capabilities**
This step of phase A focuses on understanding the capabilities within an enterprise, including the ability to develop and consume the architecture, and assessing the baseline and target capability levels. While it is not feasible to get a full grasp of all these elements given that the architecture involve many organisations, a brief analysis on the three elements mentioned above can be performed, and hold valuable insight:

   - **Ability to develop and consume the architecture:**
   Performing large IT transformations is a difficult undertaking for any organisation; with a majority of large scale IT projects coming in over budget, late or delivering less value than expected [99]. However, large scale transformation in the medical sector have happened in the past as technology progressed. This shows that, given the commitment of adequate resources, a change like the one presented in this thesis can happen.
   - **Assessing the baseline capability:**
   The baseline capability depends on the healthcare process in question. Some processes, like the process of a patient going from a GP to a specialist at the hospital are rather well supported. Others, usually processes which are less common, often times leave a lot of room for improvement. This shows that the baseline is inconsistent depending on the need of the patient, which is not desirable.
   - **Assessing the target capability:**
   In the ideal scenario, the data of a patient is accessible to a healthcare professional wherever, and whenever it is required. The only exception to this is when a patient has not given consent for their data to be shared. This should not be dependent on the healthcare process the patient is going through, however uncommon it may be. Secondly, there should be an easy way for researchers to access medical data for secondary usage, again if the patient has given consent to this.

[1] Through umbrella organisations like the Nederlandse Vereniging van Ziekenhuizen (NVZ), The Dutch GGZ or the National GP association. The full list of signatories can be found on page 2 of the IZA [97].

4. **Assess Readiness for Transformation**
   In this fifth step of phase A of the ADM one needs to assess the readiness for the enterprise to go through the architecture change. This is done by going over the readiness factors which TOGAF provides. However, given that this thesis covers a dataspace covering multiple enterprises, as well as the large amount of information required to perform the readiness analysis this is not possible within this research.

5. **Define Scope of Architecture**
   It is relevant for any project to define the scope of that which needs to be developed. This step of of phase A provides guidance on how to define the scope for the architecture. First is to determine the **breadth** of the enterprise activity (business sectors, functions, organizations, geographical areas) to be modelled. For the dataspace, the breadth of the enterprise is to develop an architecture which can represent all healthcare providers in the Twente region, focusing on the processes in which inter-organisational data exchange is realized. The second aspect of scope is determining the **depth** of the architecture, in other words, how detailed the architecture models are to be. The models in this thesis is of a level of detail, where core components of the IDS-RAM are identifiable, with some technical details included, and also detailed enough where experts can determine the validity and viability of the architecture in section 5. Lastly, the business, application and technology layers/domains are all included in the models.

6. **Develop Architecture Vision**
   While going through all elements present in this step is not feasible for this research project, picking out an important aspect is. This step states: '*This step generates the first, very high-level definitions of the baseline and target environments, from a business, information systems, and technology perspective*'. As such, this is presented here:

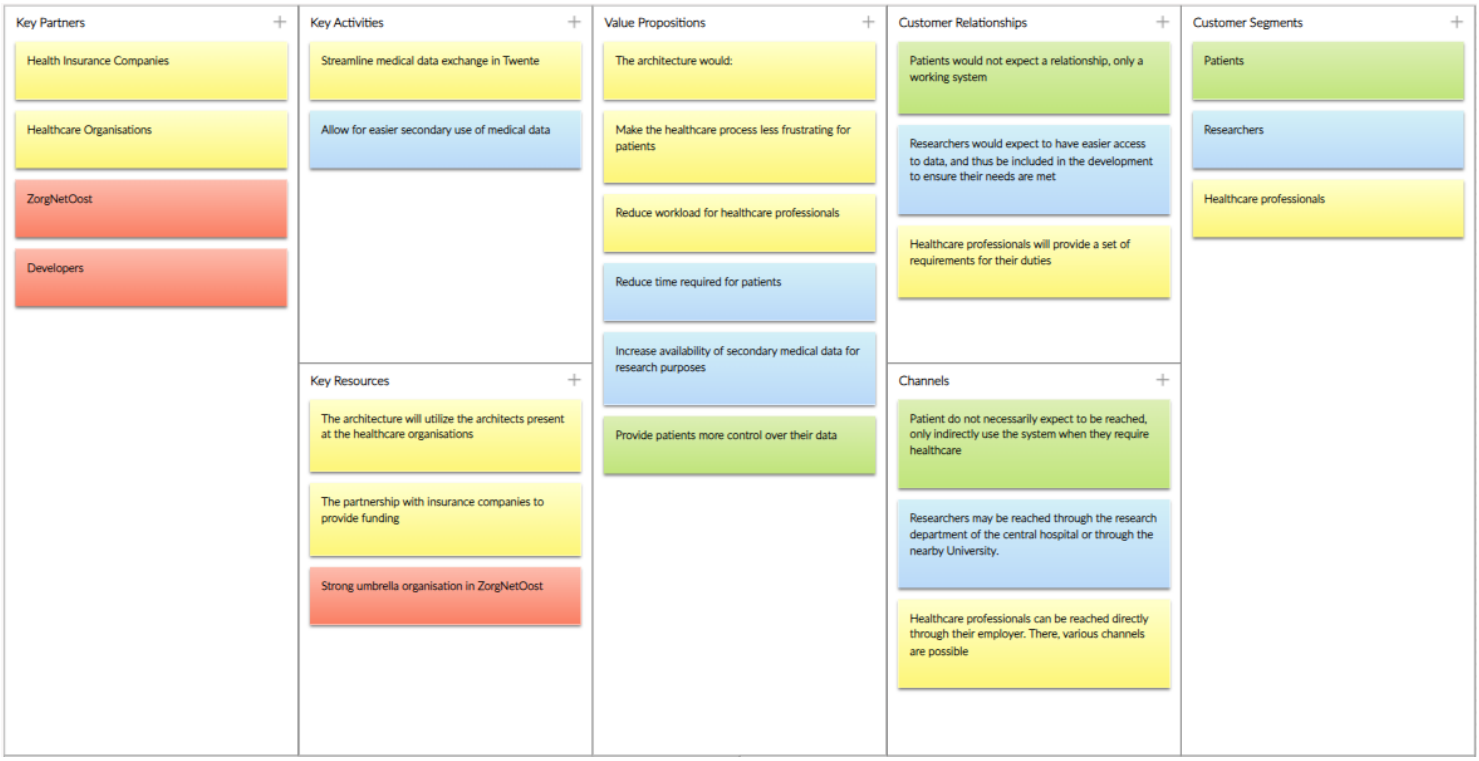
   - Baseline:
     - Business: In the current state of the business processes the exchange of medical data is highly dependent on the healthcare process a given patient is going through. In the best case scenario when going from one healthcare provider to another all information is transferred digitally already. However, in the worst case scenario paper patient files are still used.
     - Information: Currently systems are far from fully interoperable. There are standards which aid in the exchange of data, but they are not always fully agreed upon. An ecosystem should be decided upon which clarifies how data is exchanged.
     - Technology: Healthcare providers take care of their own technology stack to support the applications they choose to use.
   - Target:
     - Business: In the target state a patient should be able to go to any healthcare provider in the region, and that healthcare provider should be able to request all required medical data on the patient, assuming consent has been given.
     - Information: An ecosystem should be created which has agreed upon standards which are to be used. The standards should facilitate all needs required of the systems deployed at healthcare providers. In the presented case, the ecosystem is the dataspace.

- Technology: Similarly to the baseline, the technology is still decided by the healthcare organizations, however they should be able to support the application of the dataspace.

7. **Define the Target Architecture Value Propositions and KPIs:**
Once again, select elements of this step of phase A are relevant and feasible for this research project. These can be found here:

   - **Develop the business case for the architecture**
   To show a summarized view of the business case of this architecture, a business model canvas was created. This can be found in figure 4.1. Red represents operational elements, yellow organisational, green patient-related and blue the secondary use of data.
   - **Produce the value proposition for each of the stakeholder groupings**
   In this section, clarification is provided on what value the architecture change brings to each of the stakeholders. The stakeholders are extracted from table 3.1:
     - Healthcare professionals: More commonly have access to patient medical information, reducing workload by eliminating repeating of certain tasks.
     - Researchers: Easier access to medical data.
     - ZorgNetOostIT staff: No direct value extracted from the architecture itself, however, job security in maintaining the system could be seen as value.
     - Health Insurance Companies: More efficient healthcare processes can reduce the overall cost of healthcare, reducing claims to the insurance companies.
     - Patients: Avoiding repeatedly answering the same questions, leading to a more efficient healthcare process.
     - ZorgNetOost: The healthcare environment in their region should work better after deployment of the new system.
     - IDSA: Another operational dataspace following their reference architecture, assuming the dataspace is put into practice.
     - Developers: Similarly to the IT staff of ZorgNetOost, no direct value from the architecture change but job security in developing the system.

FIGURE 4.1: Business Model Canvas of the Architecture Change



As mentioned before, the ADM is modular to the needs of the project. Additionally, not all steps are relevant for a research project like this thesis. Below you will find each of the steps of phase A which were not discussed in detail. Along with each, a explanation is provided as to why it was omitted.

- **Establish the Architecture Project**
  The first step of phase A is the establishment of the Architecture Project. This step is in place to establish recognition for the project within the enterprise to ensure support and commitment. Given that this thesis is a research activity, with confirmed support in place by supervisors, this step requires no further attention.

- **Confirm and Elaborate Architecture Principles, including Business Principles**
  This step was not included because it is based on Architecture Principles. '*Architecture Principles define the underlying general rules and guidelines for the use and deployment of all IT resources and assets across the enterprise.*' [96]. The drafting of these principles was skipped as priority was assigned to other aspects in this thesis. Drafting these principles required gathering consensus among most stakeholders in table 3.1. This was deemed not feasible.

- **Identify the Business Transformation Risks and Mitigation Activities** There is risk in any large business IT transformation like the one described in this paper. However, due to time constraints, this step of phase A was omitted. General risks of IT projects include [100]:
  - Scope drift
  - Budget overruns
  - Timeline issues
  - External risks, related to the client
- **Develop Statement of Architecture Work; Secure Approval** Given this is a research project, it has already gained approval and thus, this step is skipped also.

4.2 Phase B - Business Architecture

The objective of phase B is to develop the target business architecture. In practice this means developing a model which describes how the enterprise operates to achieve the business goals. Much like phase A, TOGAF uses a number of steps to describe the process required to complete phase B.

1. **Select Reference Models, Viewpoints, and Tools**
   This first step of phase B is a very comprehensive one. It requires the architect to select business architecture resources, like reference models or patterns and also business architecture viewpoints like operations or management. Following those two, tools and techniques for the capture and modeling of the selected viewpoints is performed.
   Existing business architecture resources can be extremely helpful in assisting the development of new business architecture models. In table **??** below external sources utilized to develop the models in this phase are listed, along with a justification for what was used.

Source	**Utilization**
EngD Thesis D. Firdausy [88]	Thesis presents designs for specific components of a dataspace. The dataspace in the thesis is aimed at the logistics sector, however, certain elements overlap between dataspace components, regardless of sector. Due to this, thesis provides valuable information on how to design elements which are present in the healthcare dataspace as well. Specifically figure 4.14 is based on a model from this paper.
Zira (Ziekenhuis Referentie Architectuur) [47]	This reference architecture repository presents a wide variety of business process models for hospitals in the Netherlands. This thesis uses these business process models to represent the baseline business architecture (figure 4.4 to 4.11). The models are adapted to fit the focus of this thesis more appropriately.

TABLE 4.1: Phase B Architectural sources and their utilization

These sources directly presented ArchiMate models. Additional literature utilized will be cited in the explanation for each of the models where relevant.

Viewpoints discussed in this section are in table 4.2 below. The table includes what the viewpoint is, who the relevant stakeholders are, what the justification of the viewpoint is and whether it is part of the target or baseline architecture:

Viewpoint	Stakeholders	Justification	Baseline or Target
Objective/ Motivation	All except IT staff and developers	The reason for this viewpoint is to clarify to stakeholders the primary goals and objectives of the project.	Baseline
Patient Process	Patient, Healthcare providers, Healthcare professionals, Researcher	Showing the business process of a patient going through the healthcare system.	Baseline
Healthcare Provider	Healthcare Organisations, Patients, Researchers	Showing the current business process on the side of healthcare providers.	Baseline
Research process	Researchers, Patients	Showing the steps a researcher takes before gathering health data for analysis.	Target
Common standards	Developers, IT Staff, Healthcare organisations	Showing what standards have been selected to enable effective health data exchange.	Target

TABLE 4.2: Phase B Viewpoints

2. **Develop Baseline Business Architecture**
   In order to develop an adequate novel solution, it is important to first have a clear view of what the current baseline is. In this section a selection of viewpoints is designed and discussed to provide the view of the current baseline. The views included in the baseline are:

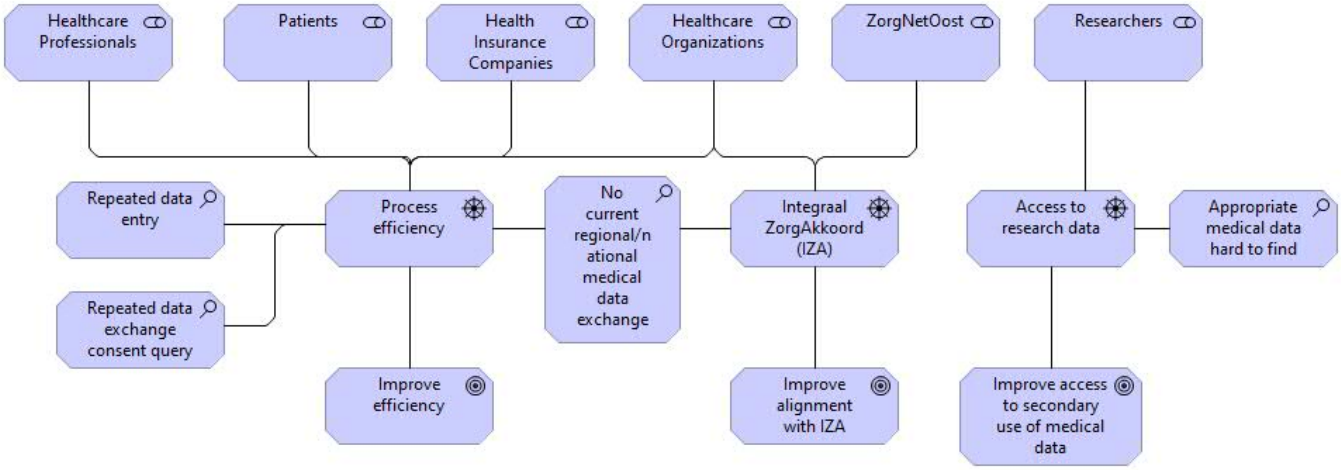
   - Stakeholder Motivation viewpoint
   - Patient Process viewpoint
   - Healthcare Provider viewpoints

   These three viewpoints focus on the the current business process involved in providing a patient with healthcare, both from the perspective of the patient and the perspective of the healthcare provider. Given that they are part of the baseline architecture, no dataspace related elements are presented here.

   Stakeholder Motivation viewpoint:
   Figure 4.2 show the stakeholder/motivation viewpoint. As mentioned, the goal of this viewpoint is to analyze the objectives and motivations behind the project and how they relate to each other. The reason for this viewpoint is to clarify to (financial) stakeholders within the operating region why a dataspace could be beneficial, and how it contributes to their objectives. The viewpoint model is best read from top to bottom. At the top all relevant stakeholders are shown, below that are the business drivers, all of which are related to observations and goals for how to improve the drivers. After the target architecture is finished an updated version could be

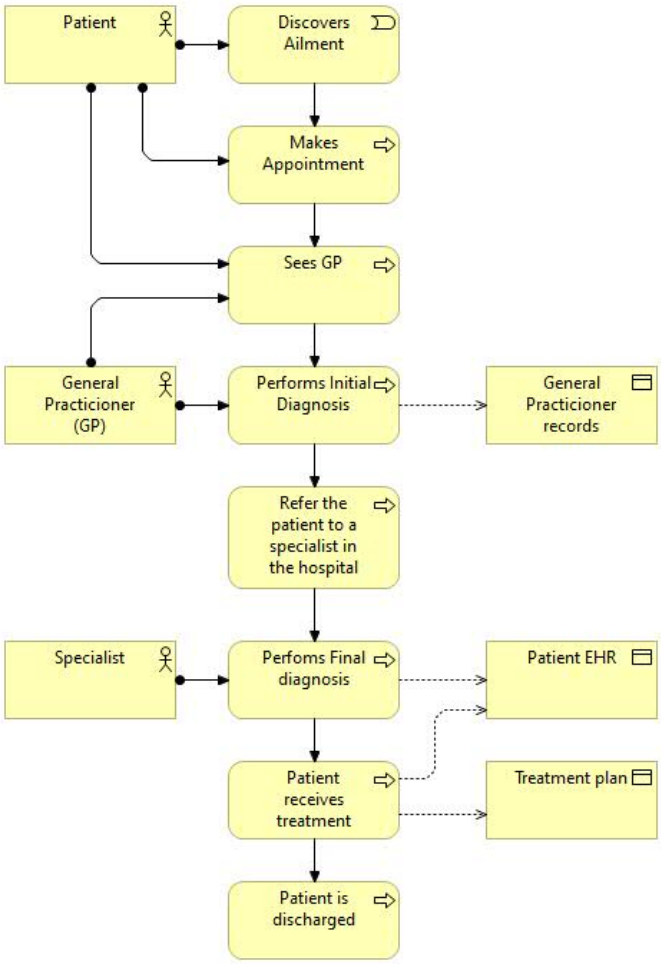
presented of this model, providing a solution to the "No current regional/national medical data exchange" element in the middle, using the dataspace.

FIGURE 4.2: Stakeholder Motivation viewpoint



Patient process viewpoint:
The first model is the patient process view. This is a very broad description as patients exist within various phases of the healthcare process. This model uses the following scenario to describe the business process as encountered by a patient. The patient notices some form of ailment and decides to go to their general practitioner (GP). The GP redirects them to a specialist at the hospital where a diagnosis is made, a treatment is selected and executed and the patient, after being treated is discharged from the hospital. Note that this model is made from the viewpoint of the patient, and thus, all processes on the healthcare providers side is not included. These processes is included in the viewpoint following the patient process viewpoint (figure 4.4). The patient process viewpoint can be found in figure 4.3.

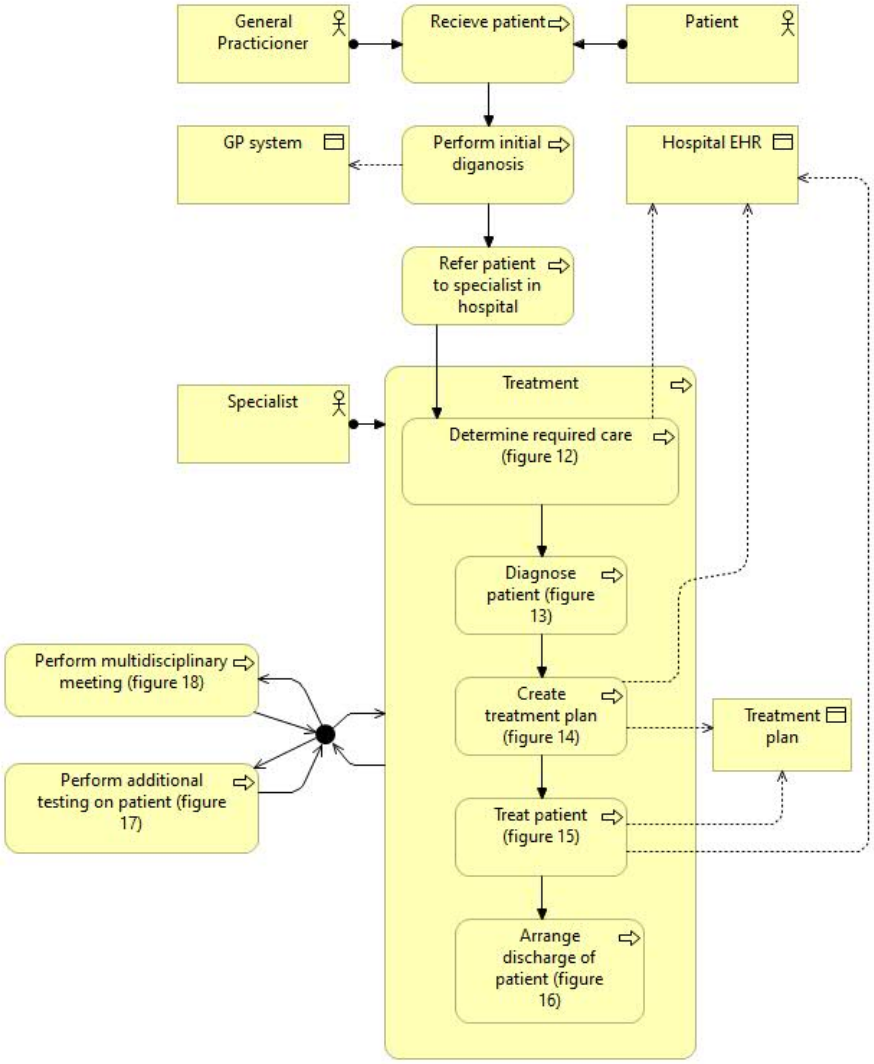
FIGURE 4.3: Patient Process viewpoint



The model starts at the top and moves down as the patient goes through the healthcare process. First the patient sees a GP who performs an initial diagnosis and logs that into the GPs electronic records. They then refer them to a specialist in the hospital who preforms further, detailed diagnosis and decides treatment. Both these elements are logged in the EHR again. Finally, after receiving treatment the patient is discharged from the hospital. Sadly, a Dutch expert source source to confirm this process was not found, however, an expert at MST did confirm that this process is correct.

Healthcare provider viewpoint:
This model serves as a parallel to the previous viewpoint (figure 4.3). It shows the same case, but from the viewpoint of the healthcare provider. As mentioned, this model draws inspiration from the reference architecture model provided by ZiRa. As the healthcare provider side is significantly more detailed, this model is divided into several smaller segments to maintain simplicity and legibility of the model. The first figure, figure 4.4 is the overview of the entire process. Subsequent models, figures 4.5 to 4.11 provide detail to the main overview model.

FIGURE 4.4: Healthcare provider viewpoint

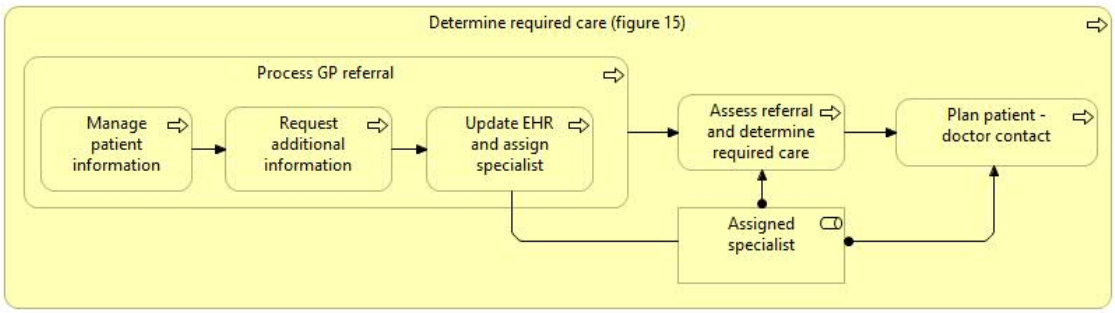


This first overview model shows the entire process. It starts with the patient being received by the GP, receiving an initial diagnosis and then being referred to the hospital specialist by the GP. At the hospital a first initial care requirement is established and the patient is then diagnosed. Following the diagnosis the treatment plan is created. This treatment plan is then used to treat the patient and if everything is successful, the discharge of the patient can be arranged. In this process various data is logged in both the patient EHR and their treatment plan. To provide an example of what goes in each, a patient EHR will contain elements like their medical history, allergies and vaccination history. A treatment plan contains elements related to one specific health condition, elements like diagnosis, goals and treatment strategy. Alongside the diagnosis, treatment plan creation and treatment of the patient a multidisciplinary meeting or additional testing can be performed.

Keep in mind that this is the baseline, when everything goes well. Often the referral of the the initial caregiver (the GP in this case) to a specialist at the hospital does not go smoothly. This is one of the problems which the dataspace aims to alleviate.

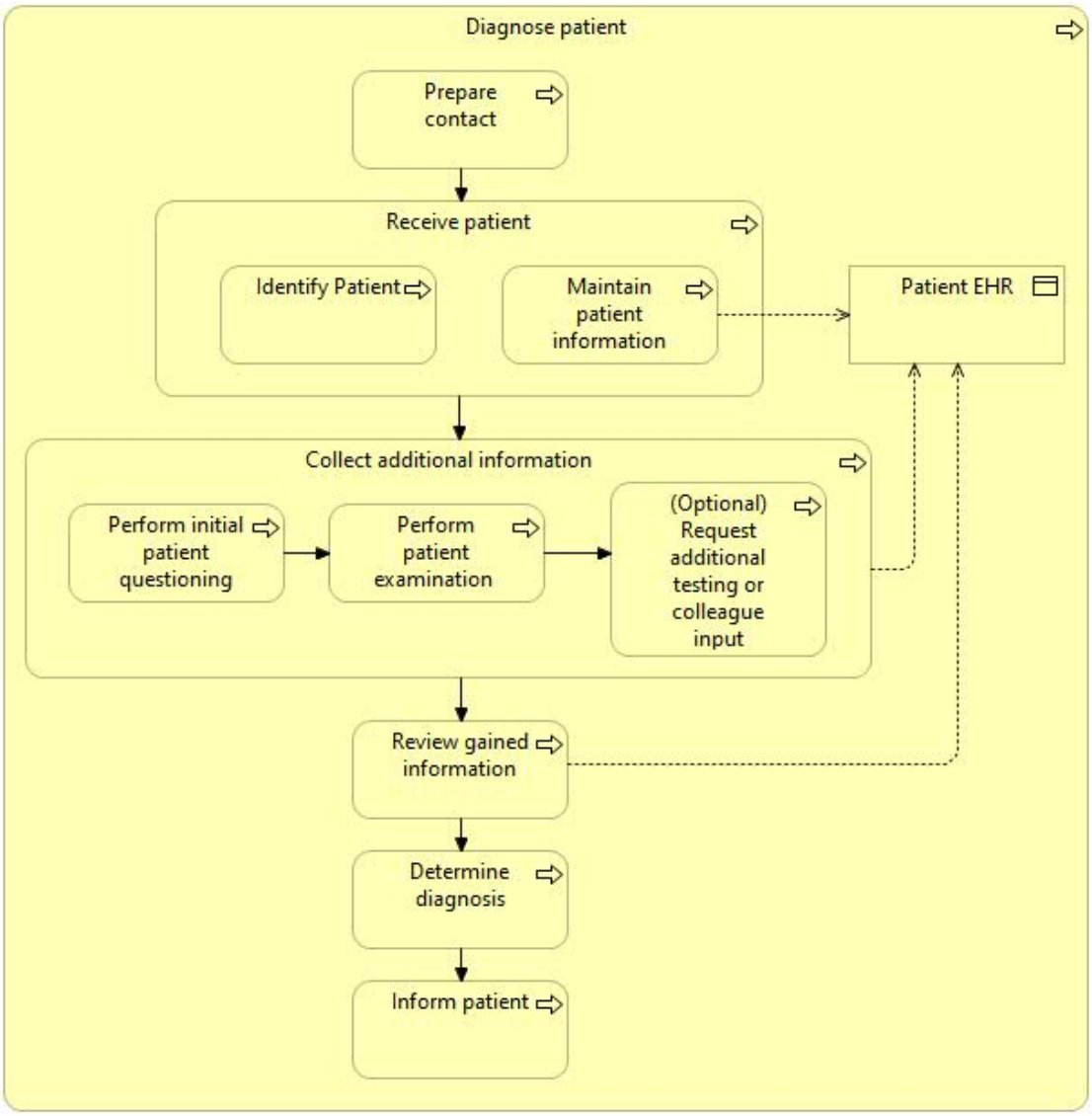
Determine required care:
Figure 4.5 shows the first sub-process of the overview model from when a patient arrives at the hospital. First the referral provided by the GP needs to be processed. This means ensuring that the referral is accepted, all patient information is logged in the hospital systems, request patient data from external sources and making sure that information is present in the patient EHR and lastly assigning a healthcare professional to the patient. This assigned specialist then assess the referral and information from the first step and plans a contact moment with the patient. When this is done, the process moves over to the next step, diagnosis (figure 4.6).

FIGURE 4.5: Determine required care



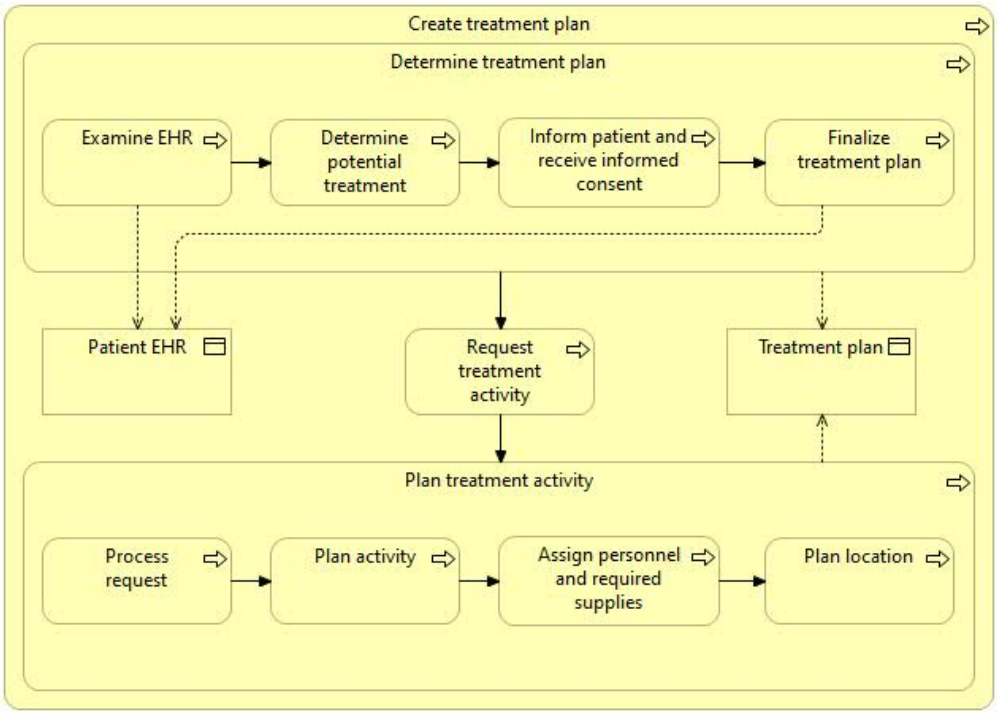
Diagnose Patient: Figure 4.6 shows the second sub-process in which the diagnosis of the patient is performed. The contact with the patient is prepared to ensure the appointment is performed efficiently. The patient is received at the appointment and identified to ensure all information on the patient is correct, if not, this is maintained. Following this, the doctor moves to collect the information required to make a diagnosis. This is done by questioning the patient about their ailment, and their prior medical history. Then the doctor performs the examination of the patient and then if required, requests extra testing or input from a different specialist. The information gathered in the previous step is then reviewed and logged in the patient EHR. Using this information a diagnosis is made and the patient informed.

FIGURE 4.6: Diagnose patient



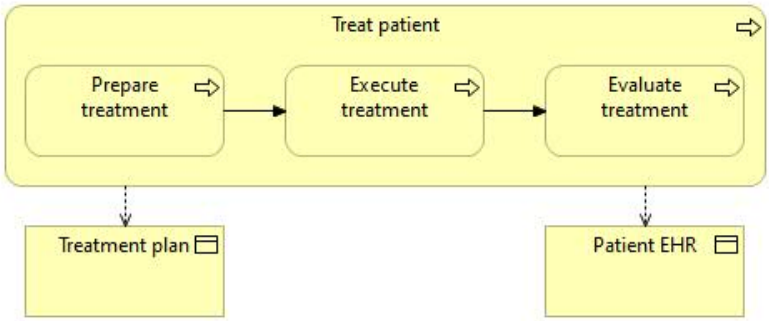
Create Treatment plan:
This next model (figure 4.7) shows the business process for creating a treatment plan for the patient. First the EHR is examined to analyze the conclusions drawn in previous steps. Then, the potential ways to treat the patient are determined. Then the patient is informed of all aspects of the treatment plan, and consent to proceed with the treatment is obtained. With that, the treatment plan is finalized. The treatment plan has some kind of treatment activities which need to be planned (an operation for example), and these have to be requested. When this is done, they can be planned. The treatment plan is also saved in the patient EHR.

FIGURE 4.7: Create treatment plan



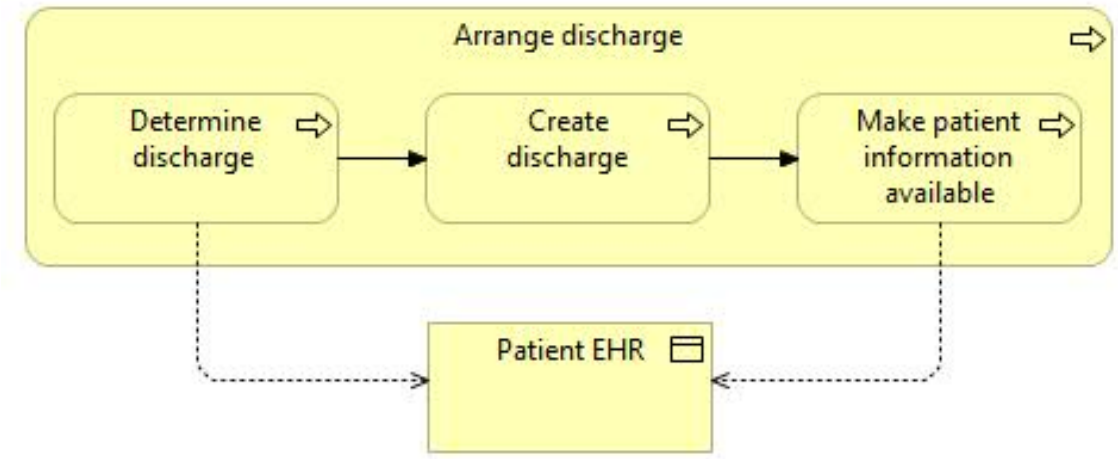
Treat patient:
This model (figure 4.8) represents the process of treating a patient. The reference model provided by ZiRa is a very comprehensive model with a variety of possible treatments. Given that this thesis focuses on the exchange of data and not the exact treatment of patients, this model was reduced to a generalized three step process, prepare, execute and evaluate treatment. This was done as the exact medical context is not relevant here to the dataspace.

FIGURE 4.8: Treat Patient



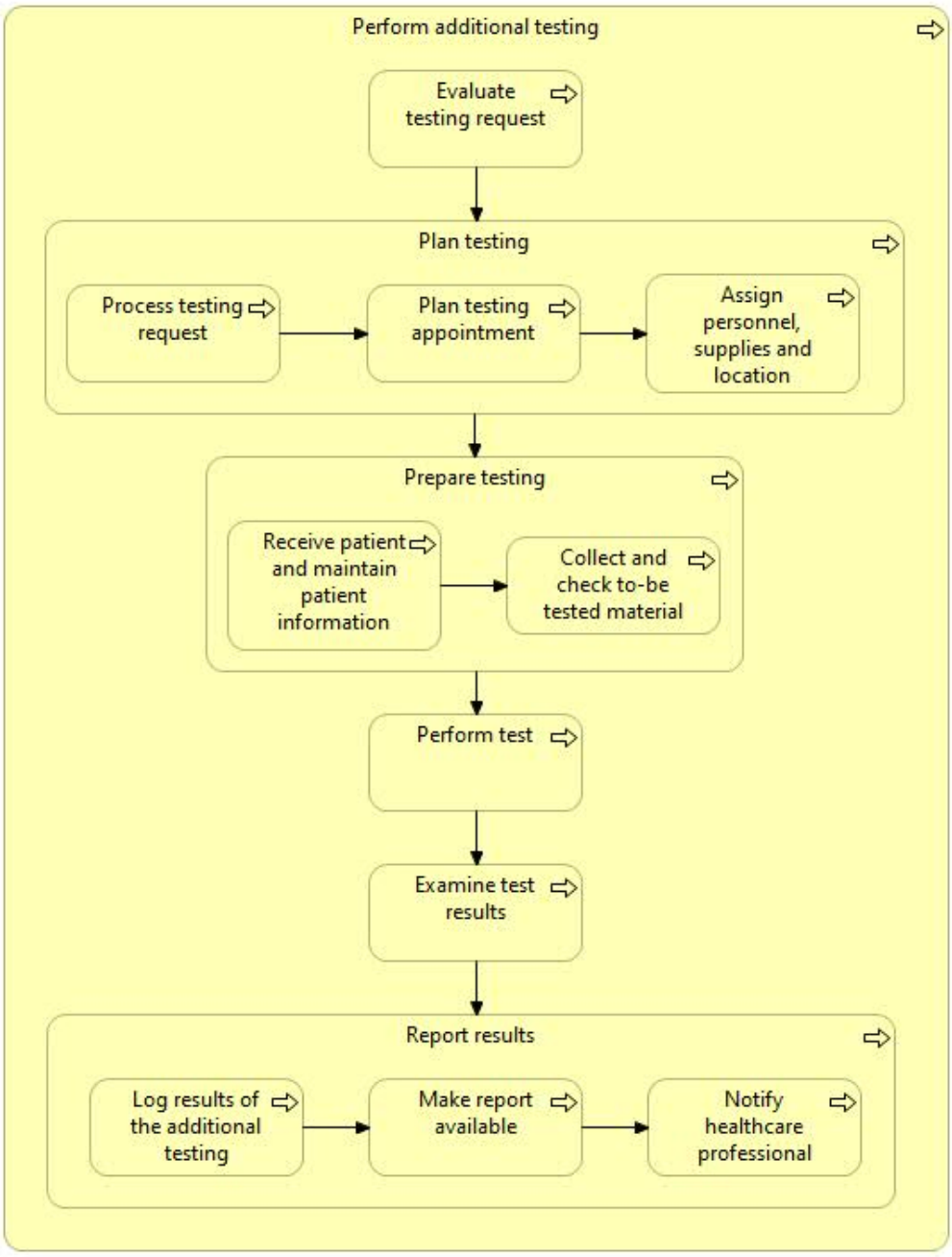
Discharge patient:
This small model (figure 4.9) shows the process for discharging a patient from the hospital after treatment is successful. The first step is determining whether the patient is fit to be discharged from care. The second is to facilitate that discharge and lastly, and the most relevant step for this thesis, is to make the patient information available for future use. The model in ZiRa also covers transferring the patient to a different care facility, making that step of crucial importance.

FIGURE 4.9: Discharge Patient



Perform additional testing:
Figure 4.10 shows the process for when a patient requires additional testing during their care journey. At the top it starts with evaluating the testing request as created by the treating specialist. This request is than processed, meaning that it is checked whether it is complete and accurate. Then the appointment is planned and the required personnel, supplies and location are assigned. When the moment for the testing has come, the patient is received and their information is checked, and if required updated. If any material needs to be collected, this is also done in this preparatory phase. Then the test is performed and the results evaluated. The results are then logged in a report, and this report is made available to the treating specialist, after which they are also notified.

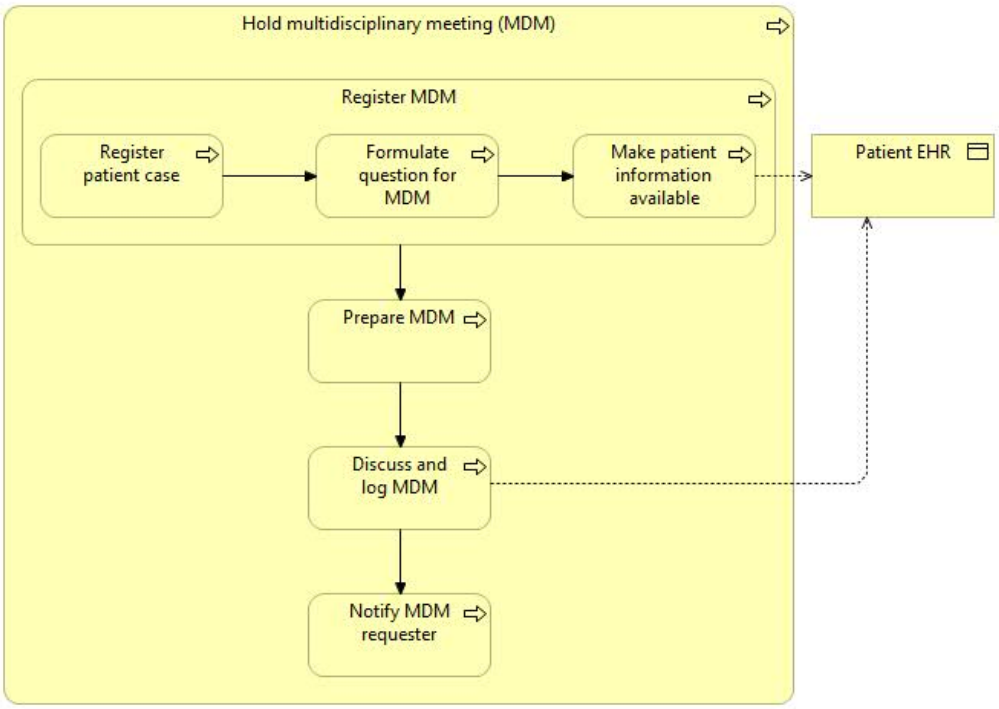
FIGURE 4.10: Perform additional testing



Hold Multidisciplinary meeting:
This last healthcare professional process model (figure 4.11) puts forth the process for a multidisciplinary meeting (MDM) to discuss a patient's case. This happens when the opinion of other doctors is needed during a patient's care journey. First the MDM needs to be registered. This involves first registering the patient's case, formulating the healthcare question to be discussed during the MDM and lastly making sure the participants in the meeting have access to the patient information to be able to give informed input. The MDM is then prepared which involves checking the

registration, adding the case to the discussion list and sharing the latest information with meeting participants. The case is then discussed and the outcome logged. Lastly, the person whom requested the MDM is notified that the outcome of the MDM has been finalized.

FIGURE 4.11: Perform additional testing



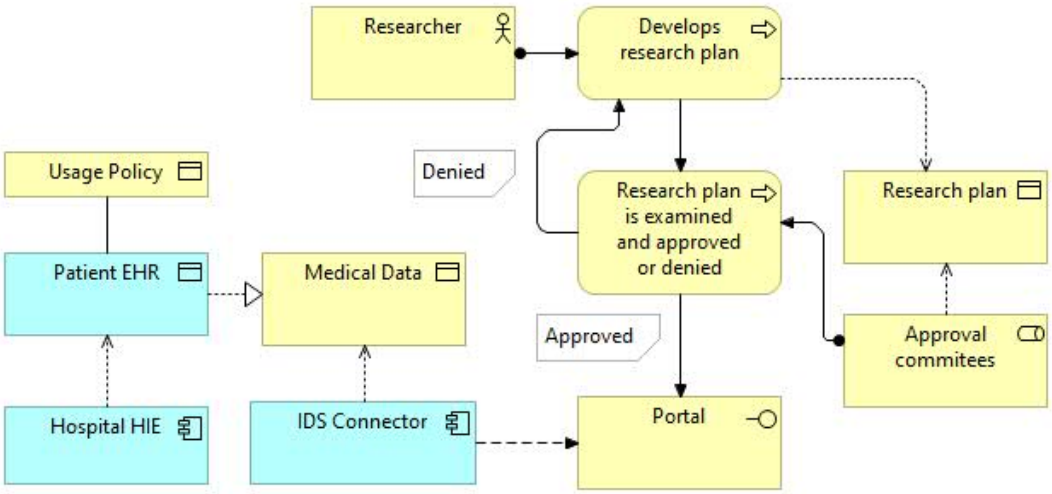
3. **Develop Target Business Architecture**
   In this step, the following viewpoints are modeled as they are part of the target business enterprise architecture:

   - Researcher viewpoint
   - Medical Standards viewpoint

   Similar to the baseline architecture, both models are shown combined with a brief explanation of the model.

Researcher process viewpoint:
This next model (figure 4.12) shows the viewpoint of the process a researcher goes through to gain access to medical data in the dataspace. Given that this thesis focuses on the exchange of data, and not the details on how a research plan gets approved, this is generalized in the model, but please note that in practice this is a rigorous process in order to ensure the privacy of the research population.

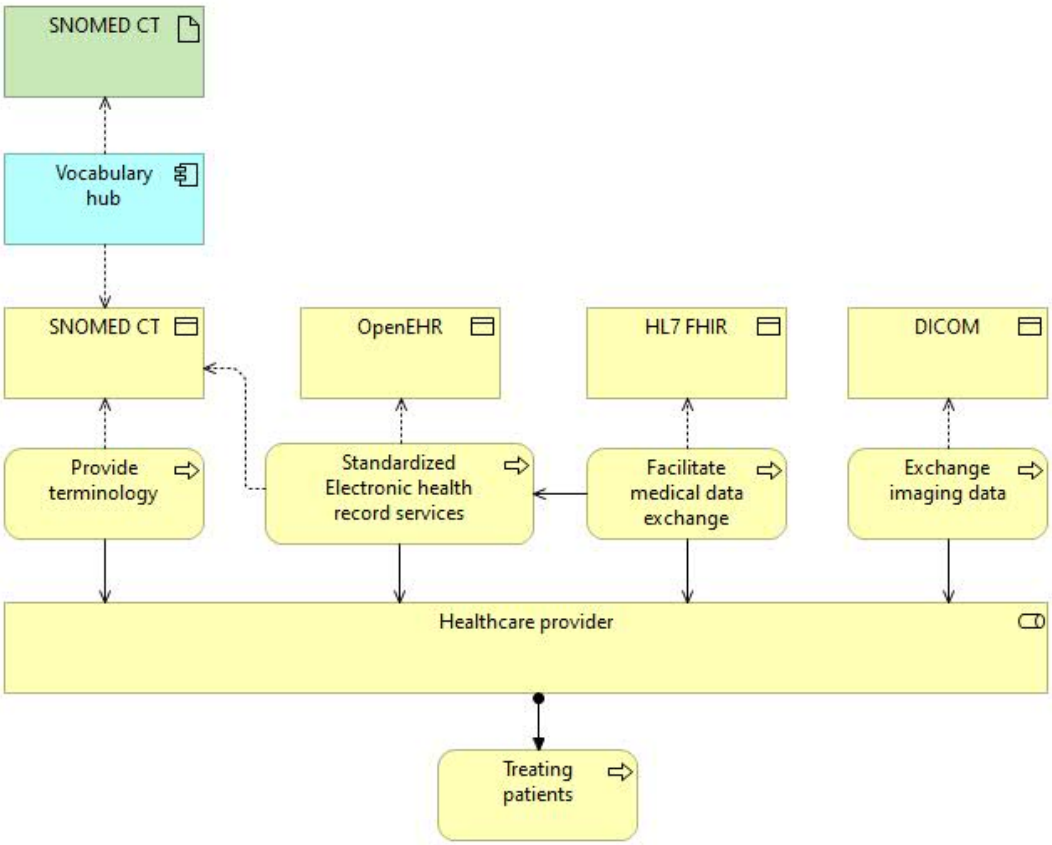
FIGURE 4.12: Researcher process viewpoint



The model starts at the top with a researcher developing a research plan, including describing the data they require. This research plan is then examined by various committees (the topic and the details of the research influence which exact committees need to provide approval). If the research plan is denied approval, then the researcher needs to go back and rewrite their plan. If it is approved, the researcher is to gain access to the dataspace (and by extension the approved medical data) through a portal which functions as a connection to a data consumer in the dataspace. This design choice was made as there needs to be a link to the dataspace and a portal allows the researcher to become a participant as the aforementioned data consumer. This portal connects to an IDS connector which in turn gets access to medical data which has been made available from patient EHRs depending on the patient usage policy.

Medical standards viewpoint:
This last model (figure 4.13) shows the viewpoint of the standards which have been highlighted as the most important standards in the dataspace. These are SNOMED CT, OpenEHR, HL7 (FHIR) and DICOM.

FIGURE 4.13: Common standards viewpoint



The model shows the four standards horizontally aligned. Each of the standards have a summarized description of their role as a business function. These functions then serve the healthcare provider shown at the bottom. A link to the dataspace is shown through the inclusion of the vocabulary hub. The four standards in the dataspace from this model are SNOMED CT, OpenEHR, HL7 FHIR and DICOM. These four are selected as the cornerstone of the dataspace for the following reasons:

- **SNOMED CT:** The SNOMED standard is a very comprehensive medical terminology system which allows various healthcare providers to have a unified language within the dataspace. This unified language significantly aids in ensuring semantic interoperability between all the participants within the dataspace. Additionally, SNOMED CT was selected not just for its popularity and comprehensive terminology system, it was also concluded during stakeholder interviews that SNOMED CT was one (of two) terminology systems which primary care medical staff desired.
- **OpenEHR:** OpenEHR is a platform for maintaining and storing electronic health records. One of its strong advantages is that it is open source, and vendor independent. It also separates the data from the applications. This allows organisation within the dataspace to update applications without altering the underlying data.
- **HL7 FHIR:** HL7 FHIR is a standard which enables exchange of medical data. It is based on RESTful-services, making it a rather lightweight way of sending and

receiving data. It offers a modern (it is still being updated regularly) and provides a lot of flexibility to the user. Additionally, similar to SNOMED CT, from stakeholder interviews it was determined that HL7 FHIR is the preferred standard for data exchange given the reasons above, and its rising popularity in general.

- **DICOM:** Is a standard for the transfer of imaging data. Imaging data in this context can come from various sources within the healthcare domain, think of X-rays, MRIs or CT scans. It coexists with FHIR for a variety of reasons. First, FHIR does not yet offer solid support for transfer of image data. Secondly, DICOM is an older standard that has become deeply ingrained in a wide variety of systems over time, making it difficult to transition away from. Given these factors, FHIR and DICOM have to co-exist in order to provide full support for the entire spectrum of medical data.

For more information on these four standards please refer to section 2.2.1.

4.3 Phase C - Information Systems Architecture

After finishing the business architecture, this chapter moves on to phase C of the TOGAF ADM cycle, the Information Systems Architectures. Similarly to the previous chapter, this is done by going through the steps as described by TOGAF where this is relevant for this thesis.

1. **Reference Models, Viewpoints, and Tools:**
   Similar to phase B, this step of TOGAF is very important to the process of developing the architecture. Reference models are selected, and decisions are made what viewpoints exactly are required for the information architecture.

   This segment of this thesis uses models produced by D. Firdausy in [101]. This paper provides comprehensive models of IDS components in ArchiMate, which is used as reference models to develop the models for the medical dataspace. As mentioned before, it focuses on the logistics sector but insights on the generic structure of certain IDS components can be applied to the healthcare dataspace too as the design of the basic components between sectors is similar.

   Next, the viewpoints need to be decided upon. This phase of TOGAF targets the information systems, which include both data and application components. First, a viewpoint based on the thesis by Firdausy is highlighted as it serves as a building block for the larger dataspace model later. After which there is a higher level viewpoint detailing the dataspace as a whole. Lastly, some models showing specific interactions within the dataspace are shown. A full list of the viewpoints of phase C can be found in table 4.3 here:

Viewpoint	**Stakeholders**	**Justification**	**Baseline or Target**
IDS Connector Architecture	Developers, ZorgNetOost IT staff, IDSA	To show in detail how the IDS connector is built in the dataspace, and what functionalities it contains.	Target
Dataspace overview	Developers, ZorgNetOost IT staff, IDSA	Provides an overview of the entire dataspace using three organisations as a placeholder. Includes the required IDS-RAM components.	Target
IT Support staff	ZorgNetOost IT staff, Patients, Healthcare professionals, Healthcare Organisations	The dataspace requires certain support functionalities to keep it operational. This viewpoint shows how those functionalities are realized.	Target
Usage Policy Creation	Healthcare Organizations, Patients	Personal privacy is a large concern regarding medical data. Given the large role usage policy creation has in this matter in a dataspace, this viewpoint is dedicated to showing the process of a patient creating their usage policy, and how it is then included in the dataspace.	Target

Table 4.3: Phase C Viewpoints

2. **Develop Baseline Information Architecture**
   The baseline of the architecture is hard to define as the large number of healthcare providers in the Twente region can all use different applications to support their business processes. These applications cover the wide range of functionalities required of these organisations. To provide an indication of the applications present, below will be a list of applications which can be found at a GP and at a hospital. First a GP's office:

   - Electronic Health/Medical Record application: As the name implies, is an in-house system to keep track of all patient EHRs.
   - Appointment scheduling: Application to schedule appointments with patients, can also be external where patients can create appointments online.
   - e-Prescribing: A tool used to prescribe medication to patients and forward the prescriptions to pharmacies where required.
   - Billing and Insurance: Tool to keep track of billing related tasks.
   - e-Consultations: Some GPs offer the option to have an online consultations, these are usually supported using a tool for this specific purpose.

   This is a selection of applications GPs use to perform their duties, more exist, but this depends heavily on what GP you ask. Many of these would still be usable in the dataspace also, as they do not relate to inter-organisational data sharing. Now, baseline hospital applications:

   - Health Information Exchange system: Enables the secure sharing of patient health information across different healthcare organizations.
   - Electronic Health Record system: Provides a digital version of a patient's paper chart that provides real-time, patient-centered records.
   - Maintenance system: Things break in a large working environment like a hospital, a maintenance application allows staff to submit tickets to have things fixed.
   - Picture archiving and communication system: System used to store and view medical images, works in collaboration with DICOM.
   - Microsoft 365: A giant suite of applications for a variety of purposes from office related tasks to authentication of IT system users.

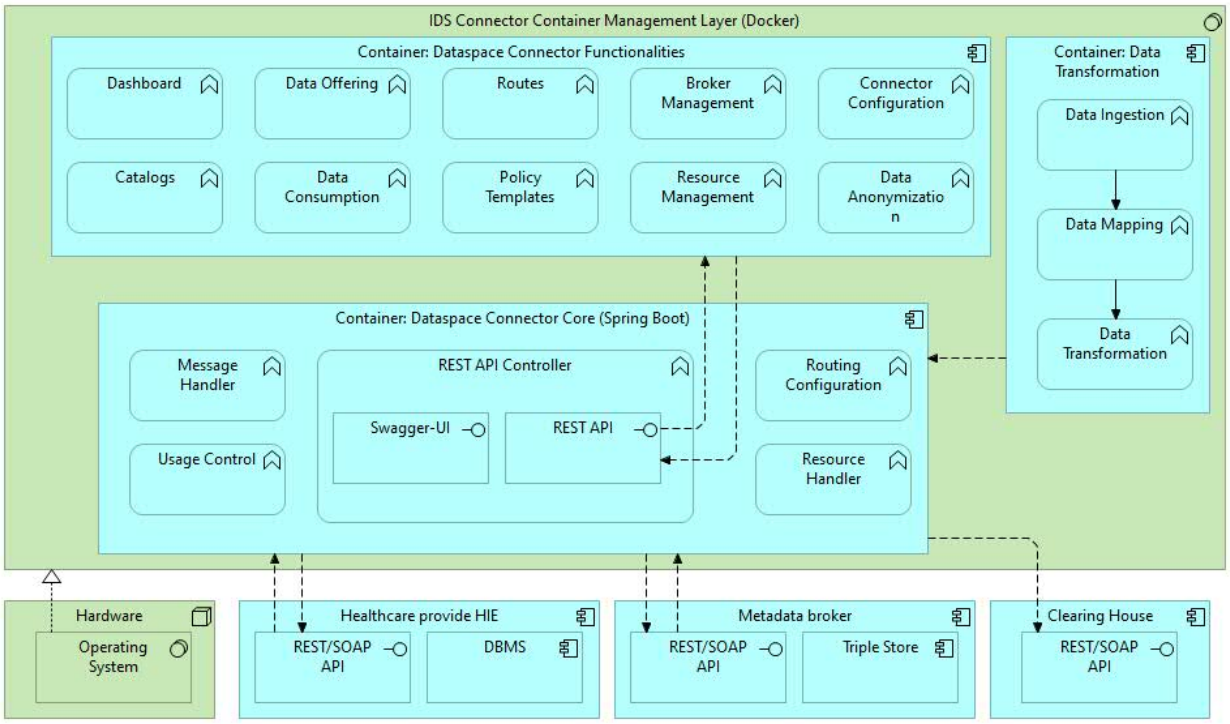
   These two lists give some indication of the very large suite of application GPs, but espcially hospitals need to function. To provide an indication, MST currently has more than 100 different applications running.

3. **Develop Target Information Architecture**
   This section will go over the target information architecture. The viewpoints included in this target architecture are:

   - IDS connector architecture viewpoint
   - Dataspace overview viewpoint
   - IT Support viewpoint
   - Usage Policy creation viewpoint

IDS connector architecture viewpoint:
The first viewpoint is the viewpoint of the IDS connector architecture. The paper mentioned in step one of phase B (Designing Essential Components For Logistics Data Spaces: Connecting Logistics interfaces, Converters, Knowledge, and Standards) provides a very comprehensive viewpoint of the design of an IDS connector, and that model can be adapted to not be specific to the logistics sector, and thus can be used in this use case too.

FIGURE 4.14: IDS connector architecture viewpoint, adapted from [101]



The paper [101] describes the model as a combined generalized version of two models created by TNO and Sovity. The connector contains several containers, each with a specialized job to execute within the IDS connector. These containers are:

- GUI
- Data App (can be multiple)
- Core
- App Management

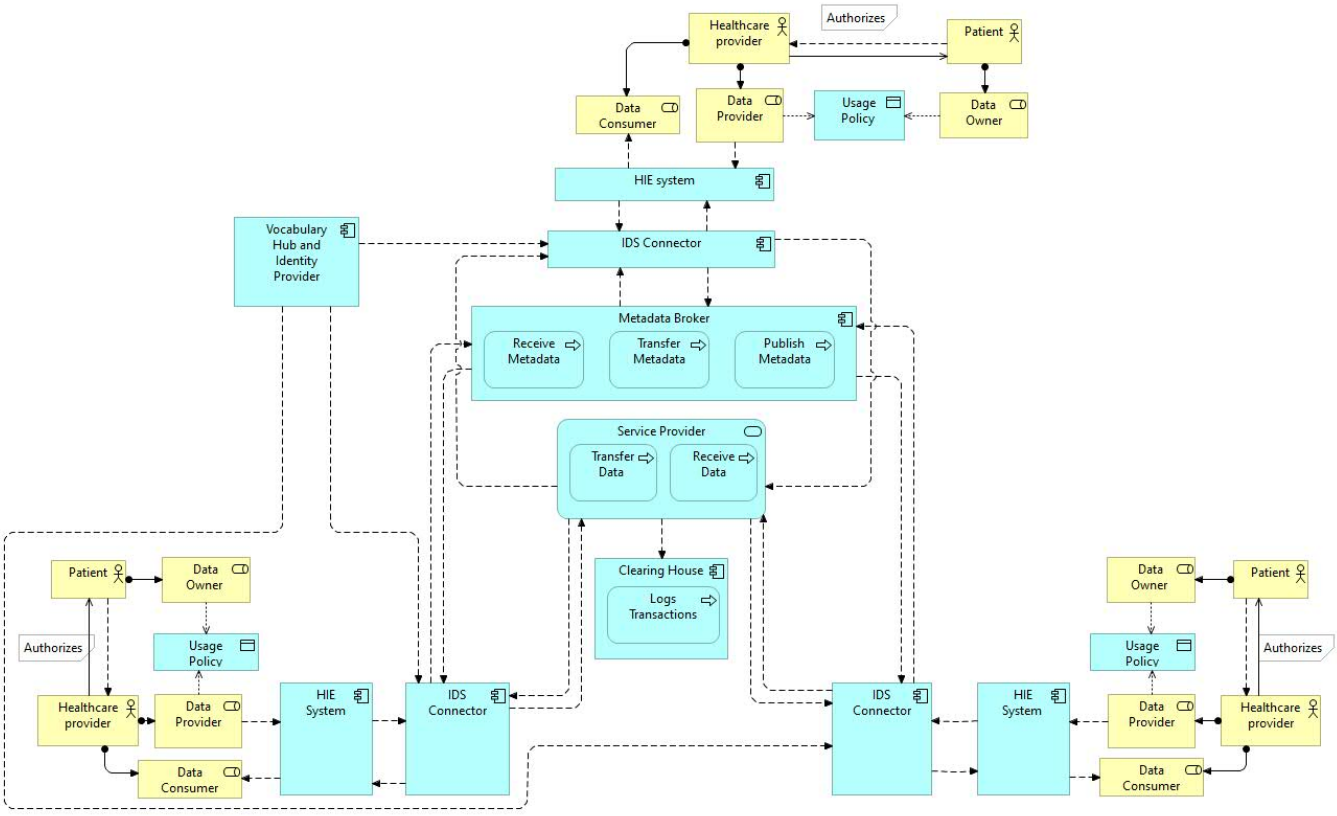
For this adapted model, the containers related to data apps were removed, as they are not as relevant for the case of a regional medical dataspace where data is mostly exchanged directly between parties and data apps are not utilized often. However, a container was added for data transformation. This addition was made to ensure that even when data is exchanged using a different format than HL7(FHIR) it can still be ingested and utilized by all parties in the dataspace. Also, the data anonymization

functionality was added to allow for data to be utilized for research without exposing personal information.

Dataspace overview viewpoint:
The next viewpoint (figure 4.15) is the dataspace overview model. This viewpoint shows, on a relatively high level, the full dataspace. The model is slightly generalized as it can be expanded to include as many healthcare providers as needed. In the model, three are included to represent these healthcare providers.

According to an IDS specification (DIN-SPEC-27070:2020-03 [102]) there are three topologies which can be used for data exchange. The first one is a peer-to-peer network where every client is equal to another and may also function as a server. In peer-to-peer networks very little central infrastructure is required. The second is the opposite of the peer-to-peer network: the client-server topology. One central server functions as a hub and all participant clients interact via that server. Evidently, a strong central infrastructure is required. Last is the hybrid topology, in which each node in the network can exist as server and a client. For this topology to function as a dataspace a broker and identity provider are required. The architecture put forth in this paper would function best using the hybrid topology as it is a good fit for the use of the IDS-RAM. Secondly, due to hospitals fulfilling the role of both client in creating data but also functioning as a regional medical hub lends itself well to designing the dataspace as a hybrid topology.

FIGURE 4.15: Overview of the generalized medical dataspace

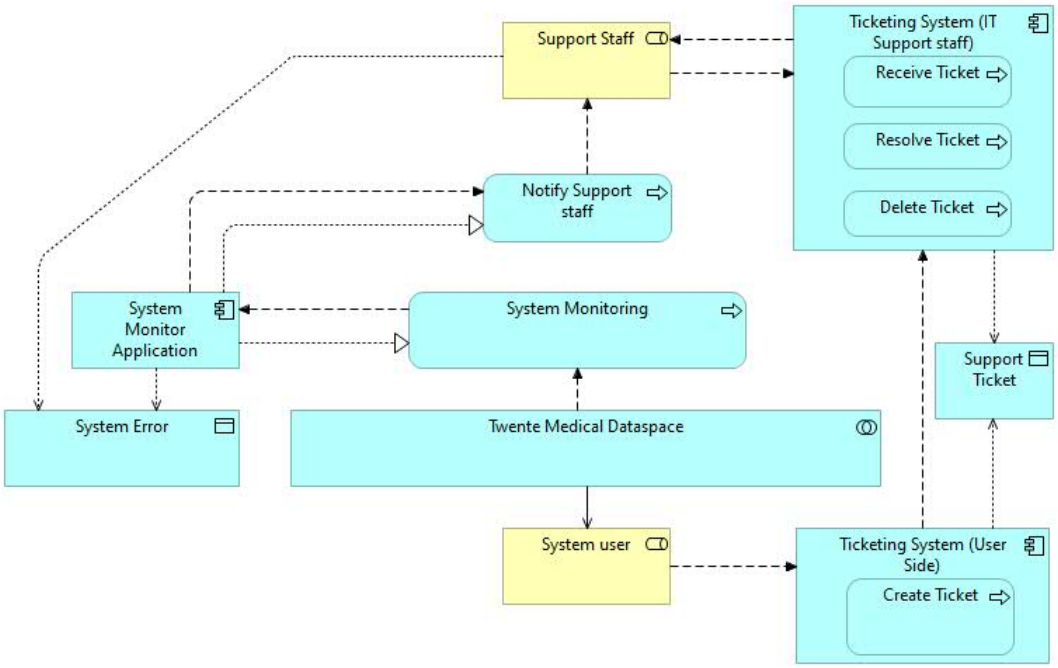


The model (figure 4.15) shows three healthcare providers within the dataspace. Each have information on patients which they are authorized to share with others following a usage policy created by the patient. Through a stakeholder interview the requirement of keeping current systems in place for nursing staff was established. This is fulfilled through having them process their information through existing Healthcare Information exchange (HIE) systems, which in turn exchange that information with an IDS connector. Through that IDS connector they can access a metadata broker to find the information they require from other healthcare providers and they provide their patients' data where authorized to the broker. When they want to send or receive data this is done through the service provider, an application service responsible for the actual data exchange between parties. Furthermore, the vocabulary hub and identity provider are both connected to all IDS connectors to provide their respective services. They are combined into one entity to improve readability of the model.

One more point to raise based on this model is the use of the terms Data Owner, Data Provider and Data consumer. These terms were introduced in section 2.3.1 and are related to the IDS-RAM. The data owner is the entity which can execute control over the data, the data consumer is the entity who uses exchanged data, and the data provider provides the data to the data consumer. In this data space, the patient can create a usage policy and through this, executes control over the data. The healthcare providers can use or provide the data, and thus can take both roles. The decision was made to use the IDS roles to communicate these responsibilities in the dataspace more clearly. Alternatively one could use the terms which are used in the GDPR [54], Data Controller, Data Processor and Data Subject.

IT Support viewpoint:
The following viewpoint, figure 4.16 describes two activities which are required of the dataspace and are related to IT support activities. These activities are system monitoring and support ticketing. Both of which are part of the requirements laid out in section 3.3.

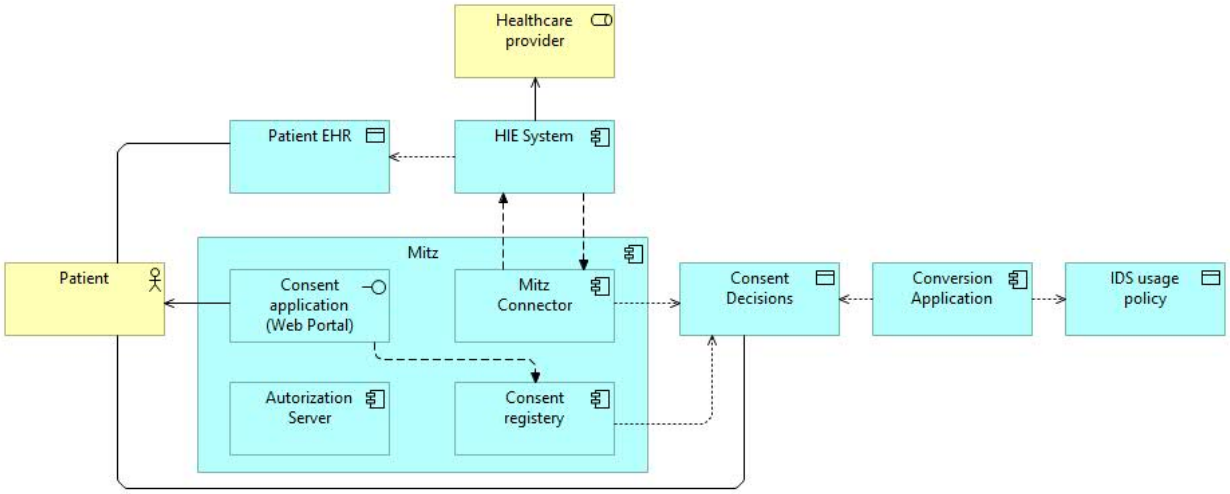
FIGURE 4.16: IT support viewpoint



The model has the dataspace in the middle. It is highly simplified as an application collaboration as the exact origin of the issues or tickets is not relevant for this model. Above the dataspace is the design of the system monitoring. A system monitoring function, monitors the dataspace for any system errors which may arise. This function is realized by the System Monitor application. This application also notifies the support staff, whom can also access the error related to the notification. Below the dataspace a system user is displayed. This can be a patient, but more likely, a stakeholder on the healthcare provider side of the dataspace. They can create a ticket, which is then send to support staff, whom then decide what to do with the ticket.

Usage Policy creation viewpoint:
The next viewpoint is one that centers around the obtaining of consent from patients to share data, and the converting of that consent into usage policies to be utilized in the dataspace. This viewpoint is of crucial importance given the privacy aspect of medical data. The viewpoint can be found in figure 4.17.

FIGURE 4.17: Architecture for developing IDS usage policy using the Dutch Mitz system



The core of the model revolves around the use of the new Mitz system. This is a system which has recently been introduced aimed at centralizing medical data sharing consent in the Netherlands.
A patient goes online and logs onto the "MijnMitz" consent application, which is a web portal where the patient goes through a number of questions to determine what this patient consents to in terms of data sharing consent. This information is then stored in the consent registry component of Mitz. Mitz uses a Mitz connector in order to communicate with healthcare provider systems. The healthcare provider can utilize a few functionalities with Mitz:

- They hold EHR data to exchange with other hospital if consented to
- They can request EHRs from other hospitals using Mitz
- They can migrate locally provided consent to Mitz
- They can localize data using Mitz
- They get notified by the Mitz system of changes in a patients consent policy

Everything mentioned above is how the current system works with healthcare providers which are part of Mitz. In order to allow this system to work within a dataspace, the consent given through Mitz needs to be converted into a IDS usage policy. In this architecture this is performed by an additional application component which takes the consent provided in Mitz, and converts it into a machine interpretable usage policy which can be utilized in the data space.

4.4 Phase D - Technology Architecture

In this section the technology architecture is discussed. The same TOGAF steps as applied in phase B and phase C are applied here as well.

1. **Reference Models, Viewpoints, and Tools:**
   Few reference models for the technology layer exist since organisations tend to create a custom solution to support the application suite they require. No papers were found which created technology layer, Archimate models in healthcare. That being said, relevant models do exist when extending the search to other sectors. First, a paper on a IDS dataspace on smart truck parking [103] can be utilized here. Lastly, the IDS-RAM holds information which can be used to design the dataspace here. The following two viewpoints are designed and described for the technology layer:

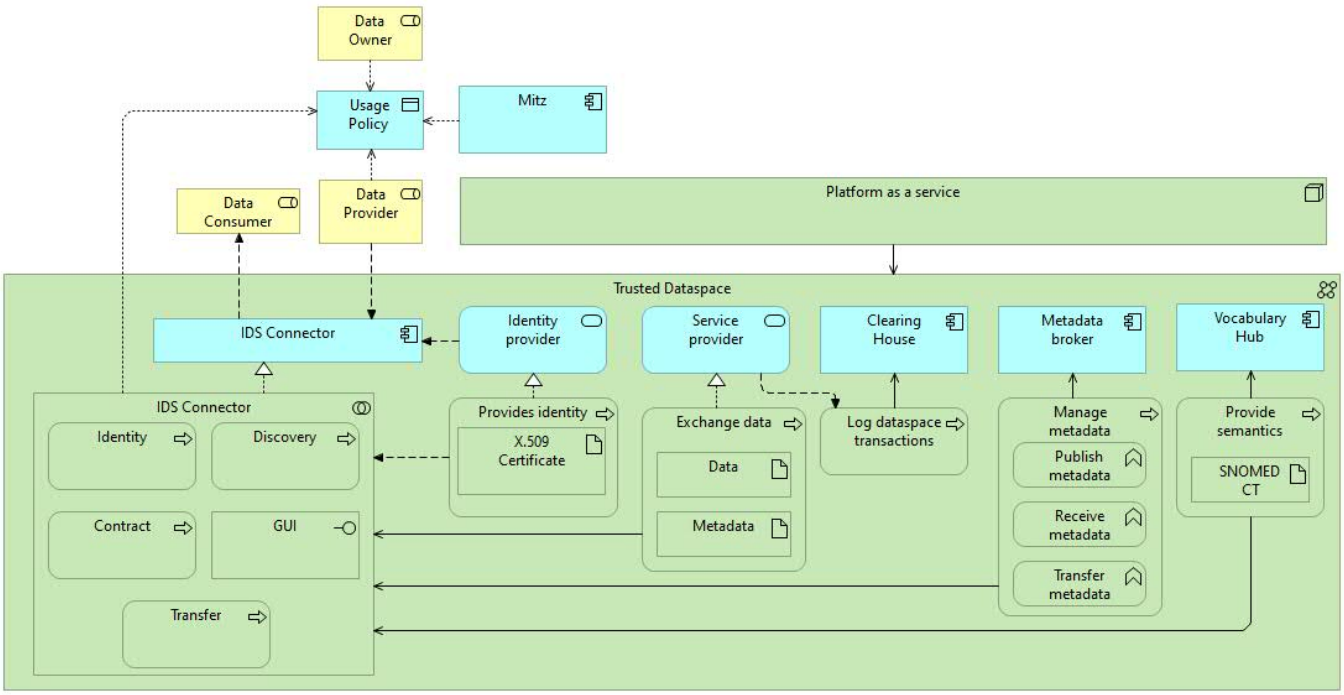
   - Dataspace technology viewpoint
   - Layered architecture viewpoint

2. **Develop Baseline Technology Architecture**
   Given the novelty of the dataspace architecture in this specific use case, the baseline technology architecture is not within scope of the research and thus omitted.

3. **Develop Target Technology Architecture**
   In this section the two viewpoints mentioned above are designed, displayed and discussed.

   Datspace Technology viewpoint:
   Figure 4.18 displays the dataspace technology viewpoint. The basis of the viewpoint is from figure 4.15. The application elements are centered horizontally in the middle. Surrounding these are the relevant technology layer elements. This model is partly based on a model from [103]. These are now discussed from left to right in the model.

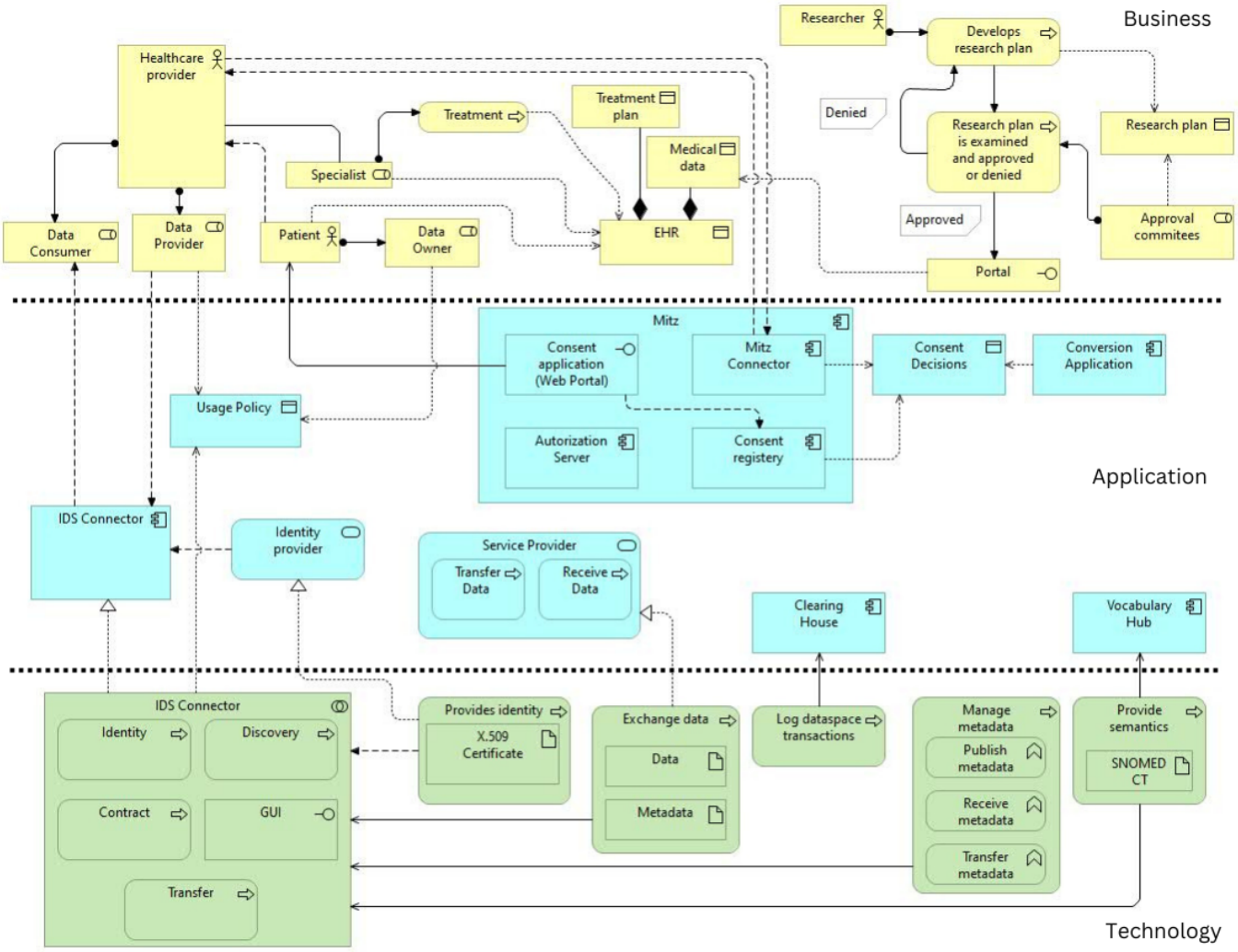
FIGURE 4.18: Dataspace technology viewpoint



First there is the large communication network which models the dataspace itself, with trusted participants. Within are the application components which were drawn from figure 4.15. The IDS connector application component is realized by a technology collaboration also called IDS Connector. This collaboration models the various technology functions collaborating to allow the application to function. Next, the identity provider is realized by the technology process "provides identity" which supplies the IDS connector with a X.509 identity certification. This is a centralized identity provider as this thesis focus on just the Twente dataspace. If in the future a network of multiple dataspaces were to be realized then using a decentralized identity provider could be beneficial. To its right is the service provider, which is realized by the "Exchange data" process. The service provider also provides the clearing house with the data exchanges which it needs to log. The meta data broker is served by the manage metadata process, which receives, publishes and transfers metadata within the dataspace. The last application component in the dataspace is the vocabulary hub. It provides the semantics within the dataspace. Given that SNOMED CT was identified as the most relevant ontology for the medical dataspace, this is included as an artifact within the technology process.
The dataspace is hosted by a generic "platform as a service" element. This was modeled this way because organisation need to choose themselves what technology structure they want to use to host the systems. Lastly, the Mitz application server was modeled to ensure comprehensive coverage of all application elements in figure 4.15. However, infrastructure configuration of Mitz is not public information and thus, the model also keeps this generic.

Layered Architecture viewpoint:
Figure 4.19 describes the entire architecture in the three layers of (from top to bottom) business, application and technology. Its aim is to give one overview of all three layers.

FIGURE 4.19: Layered viewpoint



At the top, are the most common business elements from phase B of the TOGAF ADM. The layer below features the application components which are required to make the layer above function. The layer at the bottom represents the technology layer as presented in figure 4.18. The goal of this model is to give an overview of all layers present in the architecture.

Chapter 5

Dataspace Expert Validation

In this chapter the architecture as presented in chapter 4 is validated. Additionally, this chapter includes information to answer sub questions 3.1 through 3.4. How the answers to these questions are gathered is discussed in their respective sections.

5.1 Validation of the problem identification

This first aspect of the treatment validation is the verification of the problems identified during the problem investigation in section 3.4. First, we recall the problems which were identified during the problem investigation phase:

- Lack of regional medical data exchange
- Need for interoperability and data sovereignty
- Patient privacy and regulatory compliance

For the validation of the treatment design various interview are held with industry experts (more on this in section 5.3). Included in these interviews, are stakeholders which can share their opinion on whether they agree with the identified problems. The questions can be found in Appendix B, with the conclusions from these per expert can be found in Appendix C. They described the following about the identified problems:

- Interoperability in the Twente region is currently not developed well at all, and serious improvements can be made.
- Data sovereignty is considered very important for the medical sector as privacy is an important element within data sovereignty. Slight disagreement exists on whether patients currently experience full data sovereignty.
- Generally speaking both data sovereignty and inter-system interoperability provide room for improvements, especially in terms of standardization.
- They considered patient privacy and compliance with legislation in all tasks, with a third party within the hospital verifying their work upon completion.
- None of the stakeholders questioned felt that patient privacy and regulations limit their work or have an negative impact on interoperability.

Furthermore, information on the identified problems can also be found in other sources. TwenteBeter is an organisation which represents a collaboration between various parties in the Twente region, including citizens, healthcare providers and health insurance companies [104]. All of these three are present in the stakeholder table presented in section 3.1. They have published a regional plan for improving healthcare in the region [105]. One of these aspects is improving healthcare interoperability in the region as the organisation identified issues with exchanging data between healthcare providers in Twente. This confirmed that also stakeholders outside MST experience problems regarding a lack of regional medical data exchange.

5.2 Metrics

In order to properly evaluate the architecture, certain metrics need to be determined. Given that the architecture is not implemented in this thesis, the metrics are qualitative rather than quantitative. In order to determine adequate qualitative metrics it is important to look at what the goals of the dataspace are. The dataspace needs to facilitate interoperability between both organisations and systems in the entire Twente region. In doing so, it needs to maintain patient data sovereignty and security. In line with these dataspace goals, the following metrics have been chosen and are focused on during the expert validation interviews:

- (Semantic) Interoperability: Does the architecture facilitate interoperability between organisations, systems and standards.
- Security and Privacy: Does the architecture provide appropriate access control, comply with all regulations and provide adequate security measures.
- Scalability: Is the architecture scalable to the full Twente region.

The selection of these three metrics was based on what elements were considered most important during the preliminary phase of this research. It was limited to three metrics in order to provide adequate coverage of all metrics during the interviews.

5.3 Expert validation

This section is dedicated to the validation to the presented architecture of the medical regional dataspace. This is done through a process of expert validation. The process involves presenting the architecture in a semi-structured interview, during which various segments of the viewpoints are analysed by various experts and given feedback on.

5.3.1 Use Case

In order to provide structure to the expert validation a fictional use case can be found below. This use case presents a scenario, the current problem and then the solution which the medical dataspace is to provide. The use case is utilized during the interview where the experts prefer to have the model put in a more practical setting, rather than only looking at the model itself.

Scenario: Mr. Jansen is 82 years old and lives in an elderly care facility in Enschede. He has several health issues including diabetes. At his care facility he sees a GP regularly, he also requires occasional hospital visits and receives regular wound treatment at home for

bedsores. Due to the variety of care he receives from various healthcare providers, it is of crucial importance that data can be exchanged between these various parties. In this case the GP, hospital, elderly home and pharmacy.

Problem: Between the several parties involved, medical data exchange is not always executed as effectively as one would hope. Data is not always available to healthcare professionals in the various organisations and this leads to situations where data has to either be re-collected, or requested from another organisation manually.

Solution via dataspace: The potential solution the dataspace offers is a situation where despite the medical data on Mr. Jansen being collected at various different healthcare providers, other healthcare professionals can access his data (assuming he has provided consent to this). Even healthcare providers whom in the past have never provided Mr. Jansen healthcare, should be able to access his medical information.

5.3.2 Validation Interviews

During the interview the interviewee is shown a slide containing a viewpoint, they are then asked various questions about the model. These slides and questions can be found in Appendix B.

Interview Slides and Questions:

First, the business viewpoints which were adapted from the Zira models (figure 4.3 through 4.11, page 54 to 61) are not presented to the experts. The reason for this being that these models were created by Healthcare IT specialists and only slightly adapted for the use in this thesis. Thus, to save time, these are considered accurate and are omitted from the interview.

The interviews are structured top down, where first the high level overview of the dataspace is presented, and as the interview progresses, display more detail orientated slides. Following the presentation of each slide, questions regarding that specific slide are asked. After presenting all the detailed slides, the dataspace overview model is revisited to give the interviewee an opportunity to provide additional feedback after analysing the more detailed viewpoints. The entire structure of the interview, including slides and questions can be found in Appendix B.

Interviewed experts:

Five experts were interviewed with a variety of backgrounds. The experts hold the following positions:

- Cloud Solution Architect at Microsoft
- Technical Lead at Health-RI
- Manager Data & Innovation at a consulting company
- Lead Architect at MST
- Cloud and Infra Architect at MST

Expert validation interviews conclusions:

The expert interviews were held following the template provided in Appendix B. The conclusions per model, per interview can be found in Appendix C. First conclusions are presented, followed by Points of Improvement (PoI) for the models based on the interviews. The following conclusions were drawn from all these interviews together:

- Generally the experts deemed the architecture to be **scalable**. That being said, some points of concern were raised. These concerns were mostly focused on how the dataspace would handle situations where many participants requested the same data, and the scalability of the data transformation in the IDS connector model.
- All experts pointed to **security** being a point of discussion within the architecture. Generally too little detail was included in the models and this made it hard for them to judge the security of the architecture.
- The experts were in general supportive of the use of HL7 FHIR and SNOMED CT for the dataspace. This was concluded both in the discussion during the interview, as well as during the analysis of the standards model presented during the interviews. This, together with inclusion of data transformation in the IDS connector led to positive feedback on **interoperability** as a whole.
- The method of handling **consent management** within the dataspace was generally considered fit for the purpose. However, one expert mentioned that he felt consent should be given to certain healthcare professionals, rather than whole organisations. Also, multiple experts suggested not using Mitz, but a more generic name to avoid confusion in the future.
- One element missing from the data architecture mentioned by multiple experts is a check on **data quality** before being processed by the data connector.
- While it is included in the architecture to a certain extent, the process and support for the acquiring and utilizing of **secondary data** is limited in the presented models.

Points of improvement based on validation interviews:

The conclusions discussed above provide certain points of improvement for future iterations of the architecture. To avoid having PoIs being based on the feedback of one expert, all the points below were mentioned in at least two interviews. These PoIs are listed here:

- Have the HL7 FHIR transformation occur centrally. This avoids having to make the same transformation more than once, making it more scalable, additionally you can more easily assign more resources to the data transformation if required.
- Add more detail to how the architecture achieves security overall. This can be done by adding elements to existing models, or creating a data security viewpoint.
- In the patient consent viewpoint, move away from mentioning Mitz to avoid future confusion if Mitz becomes obsolete.
- Include checks for data quality in the system.
- Connect the metadata broker to the clearing house to log transactions of meta data too.

- Provide more detail on how the dataspace supports the secondary use of data. This could be done by adding more detail to existing models, or creating a new viewpoint specifically for this purpose.

5.4 Existing solutions

Research question 3.4 centers around other existing solutions for achieving interoperability between healthcare providers. In this section the dataspace is compared to the use of the only other current solution targeted at facilitating data exchange between healthcare providers in the Netherlands.

5.4.1 CumuluZ-based solutions

Section 2.2.4 already highlighted some existing data sharing initiatives in the Netherlands. One of the discussed initiatives was CumuluZ. The aim of CumuluZ is to develop a reference architecture for a national data infrastructure. When implemented, this offers an alternative to the dataspace presented in this thesis.

Sadly, CumuluZ is currently not at a stage where they have published comprehensive information on the design of the architecture. That being said, some of the features of CumuluZ are known following discussions with MST architects.

- Both the dataspace presented in this thesis and CumuluZ have the intention of including all healthcare providers in their system.
- CumuluZ uses the data fabric concept to share data between various data sources, the dataspace uses the dataspace.
- Both the dataspace and CumuluZ use FHIR to facilitate the exchange of data.
- Both the dataspace and CumuluZ contain support for terminologies, however, how this is executed in CumuluZ is unclear based on available documentation.
- CumuluZ uses an API to communicate between the services included in the ecosystem, and the applications which implement it. The dataspace uses APIs to communicate within the connectors and between the connectors and the various components of the dataspace.
- Both the dataspace and CumuluZ feature support for the secondary use of medical data.

Architecture Elements	**CumuluZ**	**Dataspace**
Scope	All healthcare providers in NL	All healthcare providers in Twente
Data sharing concept	Data fabric	Dataspace
Exchange Format	FHIR	FHIR
Terminology system	Unclear	SNOMED CT
API usage	Between services and applications	Within connector and DS components

TABLE 5.1: Summary of Architecture Elements for CumuluZ and Dataspace

As can be seen above, there are both similarities and differences between the two systems. However, for a more comprehensive comparison more details need to be published regarding CumuluZ.

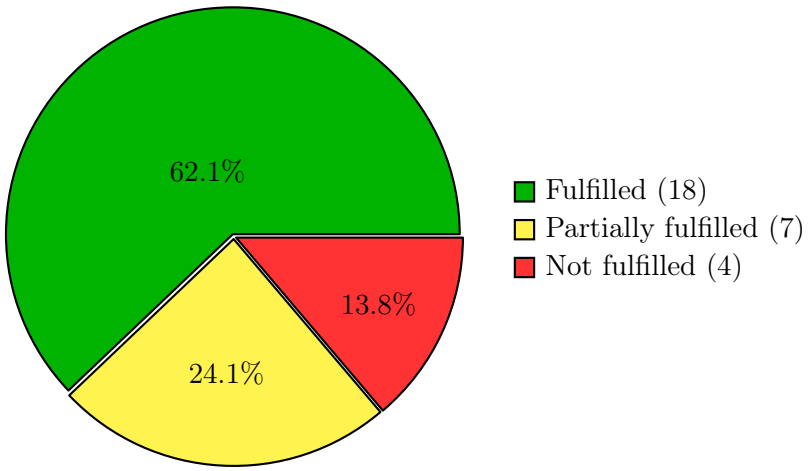
5.5 Requirements fulfillment

In section 3.3 an analysis of the 29 requirements which were collected from literature, the IDS-RAM and stakeholder interviews. This section shows the level of fulfillment for the requirements. Each requirement has one of the following tags to indicate the level of fulfillment in the architecture:

- Fulfilled: The architecture fully supports this requirement
- Partially fulfilled: Only part of the requirement is fulfilled or the requirement is not explicitly part of the design but elements in the context of the artifact could fulfill the requirement
- Not fulfilled: The architecture fails to fulfill the requirement

Below you will find a pie chart demonstrating the level of fulfillment for the requirements presented in section 3.3. In Appendix D each requirement is listed, with its level of fulfillment and an explanation.

FIGURE 5.1: Requirement Fulfillment of the Regional Medical Dataspace



5.6 Conclusion

This chapter discussed the validation of the architecture presented in chapter 4. Using the information provided in this chapter research question 3.1, 3.2, 3.3 and 3.4 can be answered:

3.1 *How do stakeholders perceive the problems identified during the problem investigation?*
In chapter 3 three main problems were identified. These were:

(a) Lack of regional medical data exchange,

(b) Need for interoperability and data sovereignty,

(c) patient privacy and regulatory compliance

Two stakeholders were interviewed regarding these problems using questions which can be found in Appendix B. Based on their answers it was confirmed that the first two problems have significant room for improvement, regarding interoperability between organisations and system. Also patient data sovereignty is not currently unanimously in a favourable state. The stakeholders did not see patient privacy and regulation as an issue. They mentioned it is simply something they have to be mindful of in their work.

3.2 *What are the metrics for evaluating appropriate data exchange for a regional medical data hub?*
Due to the architecture not being implemented in a practical setting it was not possible to set and evaluate quantitative metrics. Due to this, the goals of the dataspace were examined and qualitative metrics were determined based on the goals of the dataspace. These metrics were then used during the expert validation interviews to aid in evaluating the architecture. These metrics are:

- (Semantic) interoperability
- Security and Privacy
- Scalability

3.3 *Following an analysis of the architecture, could the architecture be an appropriate solution for medical data sharing in the Twente region?*
The answer to this question is based on the conclusions drawn from the expert validation interviews. They centered around various characteristics which can make the dataspace an appropriate solution for medical data sharing in Twente. Generally speaking the dataspace architecture received favorable feedback regarding scalability, interoperability (both between organisations and between systems) and data sovereignty. However, points of improvement were raised regarding data quality and security. Assuming that the architecture is improved following the points of improvement raised by the experts, the architecture can be considered an appropriate solution for medical data sharing in the Twente region.

3.4 *How does the architecture compare to existing data exchange solutions?*
In the Netherlands the only other data sharing initiative which has the same goal of ensuring interoperability between various healthcare providers is CumuluZ. This reference architecture provides an alternative approach to the one presented in this thesis. While similarities exist between both CumuluZ and the presented dataspace (like providing access to secondary usage data), there are also distinct differences between the two. For example, the dataspace obviously uses the dataspace concept to connect data sources, while Cumuluz opts for using a data fabric. Sadly, CumuluZ does not yet provide comprehensive documentation on their architecture. This thus leaves the answer to this question somewhat open.

Chapter 6

Discussion

During this discussion various topics related to the presented thesis are discussed. During the discussion a range of topics relevant to the findings of the thesis are discussed.

6.1 Implications for practice

Primarily, the implications for practice are relevant for three groups. First, for people working as an architect in the healthcare sector. The thesis provides an option for architects which may not have been considered in the first place. In the future, when interoperability within a certain region needs to be improved, they can consult this thesis and determine whether a dataspace could be a solution to their interoperability problems. This is mostly the case when FAIRification of the data is important, while improving interoperability between a large number of healthcare providers.

Second, the International DataSpaces Assosciation (IDSA) might have an interest in this thesis as well. While there are various dataspaces following the principles of the IDS-RAM, none of them are aimed at improving interoperability within various primary care organisations. Having this thesis to add to the growing body of literature surrounding the IDS-RAM might aid in making it more popular.

Lastly, as mentioned before in this thesis, the healthcare sector signed the Integraal Zorg Akkoord (IZA). This agreement ensures that healthcare in the Netherlands has to improve their inter-organisational interoperability. This thesis provides another avenue to achieve this increased state of interoperability. Additionally, achieving the interoperability required by the IZA is in turn also a step towards achieving interoperability on a European level as laid out in the European Health Data Space plans. So this thesis could prove useful for achieving legislative goals on a regional, national and continental level.

6.2 Relation to existing dataspaces

In section 2.1.2 various dataspaces were discussed. All these dataspaces, like the one presented in this thesis, followed the principles of the IDS-RAM. However, one large difference exists between the dataspaces discussed there, and the architecture presented in this thesis overall. All the medical dataspaces discussed in section 2.1.2 did not focus on facilitating data exchange for primary care. Most were related to the secondary, research use of data, while others focused on increasing data sovereignty to patients (for more information please refer to section 2.1.2). Based on the literature review of those dataspaces, the data

space presented in this paper is the only dataspace utilizing the IDS-RAM while focusing on increasing interoperability of medical data for primary care.

6.3 Standards version control

In this thesis, a group of standards was selected and discussed comprehensively as the backbone of interoperability in the dataspace. One element which was not discussed previously is the consideration of the version control within the selected standards.

Standardization organisations regularly release updates to their standards in order to ensure that they stay current with regards to developments in the healthcare sector. This means that, organisations using these standards have to update their systems in accordance with the standards to avoid losing support or important features. This leads to a problem where organisations are operating using different standards as not all organisations will update their systems at the exact same time. This problem is well known in API management where versioning is a task which is taken into consideration at all times.

This could also lead to a limitation in the dataspace as it has clearly defined standards to use, but no strategy as it comes to standards versioning.

6.4 Specificity to the Twente region

So far, this thesis has exclusively focused on applying a dataspace to the Twente region. This section explores whether it would be possible to change it to a different region, and what the consequences would be.

The thesis is focused on Twente because the research was performed here, and the stakeholder which could be interviewed are working in the Twente region. Due to this, requirements and current systems could only be based on the Twente region. Given these facts, the decision was made to only write about the Twente region specifically as it could not be guaranteed that the results would also apply to other regions of the Netherlands, or even the world. However, the results are mostly based on concepts which could also work in other regions, especially in the Netherlands as in general, healthcare processes in the Netherlands are the same independent of the region one finds themselves in.

When being applied to a different region some elements needs to be taken into consideration. First, during this thesis the ZorgNetOost stakeholder has a central role in coordinating the various healthcare providers in Twente. They fill the role of a "Regional Samenwerkings Organisatie" (Regional Collaboration Organisation, RCO). If the same research is to be applied to a different region, then they would have to find another party which could fill the role of an RCO in that region. Secondly, new stakeholder requirement interviews need to be performed to ensure that the requirements which were present in Twente, also apply to the new region the dataspace could be applied to.

Other than the elements above, no limitations to applying the dataspace to another region in the Netherlands were identified.

6.5 FHIR RDF and SNOMED CT Ontology

HL7 FHIR also offers a Resource Description Framework (RDF) which represents FHIR data as linked data in that RDF format. It can be combined with standardized ontologies to improve semantic interoperability in the data which FHIR exchanges. Future research

could focus on combining this FHIR RDF with the ontology of SNOMED CT. Doing so could lead to a unified semantic model which this research could show to be a valuable insight.

6.6 Personal Health Train

A recent development in the field of medical data analysis is that of the Personal Health Train (PHT). The PHT is a concept which enables secure and decentralized analysis of health data [106], and thus is a concept most closely related to the secondary use of data.

The primary idea behind the approach is that the data always stays at the data owner, and thus stays very secure. Given that the data stays with the data owner at all times, this also means that any analysis of this data has to be performed at the data owner. To facilitate this, the personal health train was introduced. The concept is called a personal health train because the train, which represents the data analysis algorithms travels to a data source, also known as a station. This decentralized approach to data analysis provides strong privacy protection as the data is never transmitted, only the results of the analysis are returned to the sender of the train.

Given the strong privacy protection it offers, the PHT is well suited for analyzing and processing of health data. Currently the offering of access to medical data for research is through a portal as presented in figure 4.12. Access to this portal and the medical data is only granted after rigorous screening by various committees. The PHT could provide an opportunity where privacy of the medical data is guaranteed, allowing for more research to be allowed to access the medical data through PHTs. That being said, no publications have focused on applying the PHT to a dataspace so far.

6.7 Success Factors of Implementation

Any large IT project has factors which can aid or harm the successful of its implementation. Those will be briefly discussed in this section.

- **Executing Points of Improvement:** Chapter 5 concluded with a list of points of improvement to the dataspace architecture. All these arguments represent clear improvements to the architecture and ensuring these are added to it would help the chances of the dataspace proving successful.
- **Adoption Rate:** In order to fully eliminate the need for manual data re-entry every healthcare provider in Twente would have to become a participant within the dataspace. The chances of this happening if the dataspace was ever to be put into practice are helped by the participation of ZorgNetOost, which represents many healthcare providers in the region. Still however, missing participants would hurt the project significantly.
- **Effective governance:** Many participants also means that there are many opinions on how governance of the dataspace should be handled. Strong communication and collaboration between healthcare organisations should aid in achieving this effective governance, and once again, ZorgNetOost plays a large role in this.
- **Patient Trust:** Patients need to be well informed about the dataspace before it being put into practice. If patients do not trust its effectiveness or privacy protection

many would most likely create a usage policy which does not allow for data exchange between many organisations. Doing so would result in a significant decrease of the effectiveness of the dataspace as this would prevent the exchange of data at this level.

Chapter 7

Conclusion

7.1 Limitations

This section will discuss potential constraints or weaknesses of the study which could limit its impact.

7.1.1 Limited Scope of Validation

The expert validation could prove to be a limiting factor. The experts interviewed are highly skilled in their domain, with each having comprehensive knowledge on their domain within information architectures. That being said, the dataspace concept, and specifically the IDS-RAM was not known to the experts and was required to be explained prior to the interviews.

Secondly, five experts were interviewed for this thesis. While an adequate amount to draw relevant conclusions regarding the presented architecture, more experts interviews would provide more perspectives and make it easier to find consistent conclusions which return more than once.

Lastly, the some of the interviews were cut short due to time constraints of the expert. This led to some interviews not containing every model. This limits the effectiveness of the conclusions regarding that model as a whole.

7.1.2 Continuing Developments

As with many things related to IT, dataspaces are continuously in development. To provide an indication of this, the Scopus database currently holds 119 papers published in 2024 alone with dataspace(s) in the title. This shows how quickly the field is moving. Due to this a limitation of the paper is the time in which it was written, with the quick development of the dataspace technology it could be obsolete relatively quickly.

Additionally, the same holds for various aspects within the dataspace. HL7 FHIR, OpenEHR and SNOMED CT are all standards which still receive updates regularly and thus change over time. These changes could also have an effect on how quickly the presented architecture is outdated.

Lastly, regulations regarding medical data is always subject to change, especially following large developments. The dataspace could require reworking following major regulatory changes.

7.1.3 Emphasis on Technical Aspects Over Social and Organizational Factors

This thesis focuses mostly on the technological aspects of the dataspace, that is the design of the various element involved in the makeup of the system. Parallel to this exists the social and organisational consideration of the system. Elements like technology adoption, staff training requirements and governance agreements also play an important role in the success of the sytem. The paper however, does not cover this aspect extensively.

7.1.4 Practical implementation

This thesis does not describe the implementation of the dataspace architecture in practice, limiting it from performing quantitative analysis of the system. The solution to this limitation is discussed in the following section, future work.

7.2 Future Work

Crucial to any discussion is the inclusion of a section on future work. In this section a number of potential future avenues of research are suggested to further develop what has been written. Given the content of this thesis the future work section will include the inclusion of the point of improvement, applying the dataspace on a more practical level, using the personal health train to support the use of secondary data and lastly, the region specificity of the dataspace presented in this thesis.

7.2.1 Working on Points of Improvement

This section explores what points of improvements uncovered during the expert validation can be added to the architecture. Additionally, more expert feedback was provided outside the expert validation interviews, this will also be discussed in this section.

Expert Validation Points of Improvement:

First the PoIs from the expert validation interviews will be discussed, and how they could be applied to the models to improve them in the future:

- The first point of improvement relates to how the transformation of other standards to HL7 FHIR is performed. In the current design, the connector features a functionality to transform incoming data into HL7 FHIR before further processing. Future research could analyze what other options are available and what is the most efficient method of doing so.
- The second point focused on the security elements in the models. According to the experts, the current models featured too little information on how security is ensured within the dataspace. Research in the future could find out what is the most efficient way to secure the processes in the dataspace and add to the models in that way.
- The next point suggested changing Mitz to a more generic name to avoid confusion in the future if Mitz becomes obsolete. This is a simple change which could be implemented immediately. Keep in mind, the current design is based on the Mitz architecture so it may take a bit of research to determine whether other architecture designs are better choices.

- Data quality checking in the dataspace was the topic of the fourth PoI. No definitive data quality checks are currently present in the architecture. Some data quality checks are present in most HIE system, but having a dedicated check somewhere in the architecture would be a valuable addition. Future research could determine what would be the best method to include this into the architecture.
- The last PoI had to do with the secondary use of data. The interviews highlighted that while the elements which were related to research were correct, there was not enough depth in the presented models. In the future a research could be performed exclusively on secondary medical data in a dataspace context to discover what options there are, and what would be the best way to include research data into the dataspace.

Additional viewpoints following external feedback:

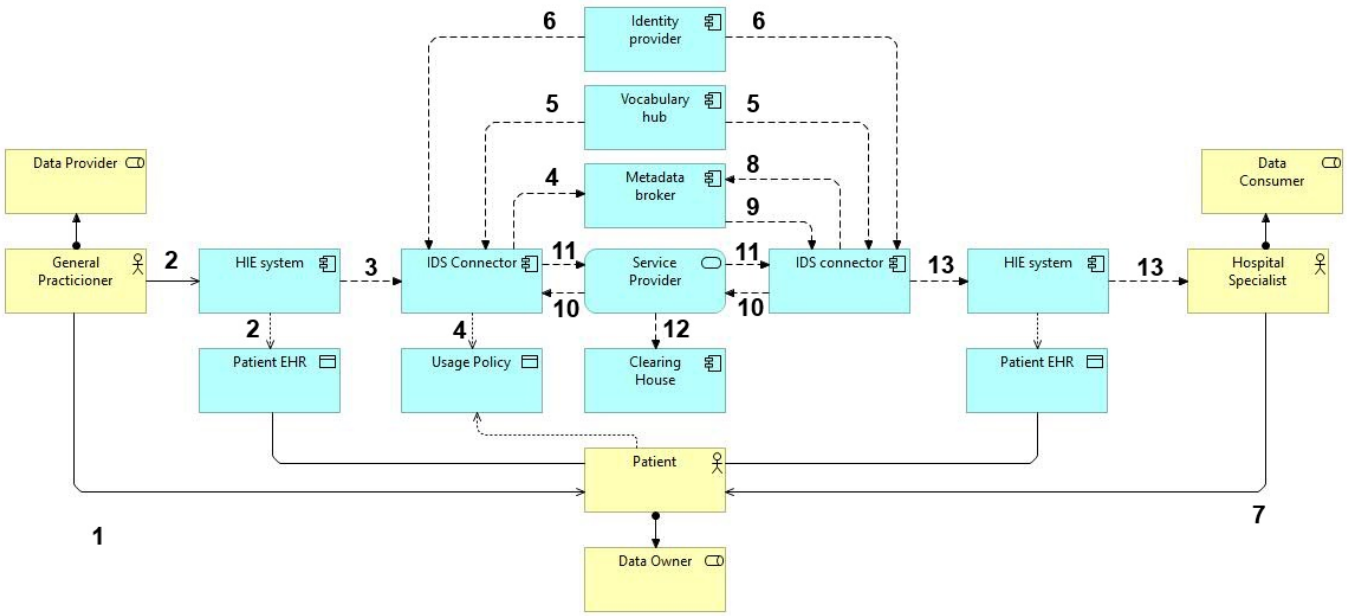
Following the feedback after the expert validation interviews, an additional model was required. Also, the standards model required significant revising. Both these viewpoints are presented here.

Annotated Dataspace process viewpoint:
The additional feedback provided following the expert interviews showed a desire for a model of the dataspace, displaying the dataspace given a specific use case. The use case is similar to figure 4.3 where a person goes to a general practitioner, and then is sent to a specialist at the hopsital. The model demonstrates how this is done using the dataspace. The viewpoint can be found in figure 7.1. The numbering was added to provide clarification to the process:

1. Patient is seen by the General Practitioner.
2. The GP reports their findings in their Healthcare Information Exchange (HIE) system, logging it in their Patient EHR.
3. The HIE system communicates the changes to the IDS Connector.
4. The IDS connector sends the metadata of the patient information to the Metadata broker according to the patient's usage policy.
5. Where required the medical terminology is provided by the vocabulary hub, it uses SNOMED CT in this dataspace.
6. The IDS connector has a trusted identity provided by the identity provider.
7. After a referral, the patient is sent to a hospital specialist, who sees them next.
8. The hospital specialist needs the new information on the patient and their IDS connector queries the metadata broker to find the required data.
9. The metadata broker sends back the requested metadata, allowing the connector to request the GP IDS connector to send over the patient information.
10. Through the service provider, the hospital connector requests the required data from the GP connector.
11. This information is then send to the hospital connector.
12. The data transaction is logged by the clearing house.

13. The hospital connector then relays the information to the HIE system, allowing the hospital specialist to use it during the visit of the patient.

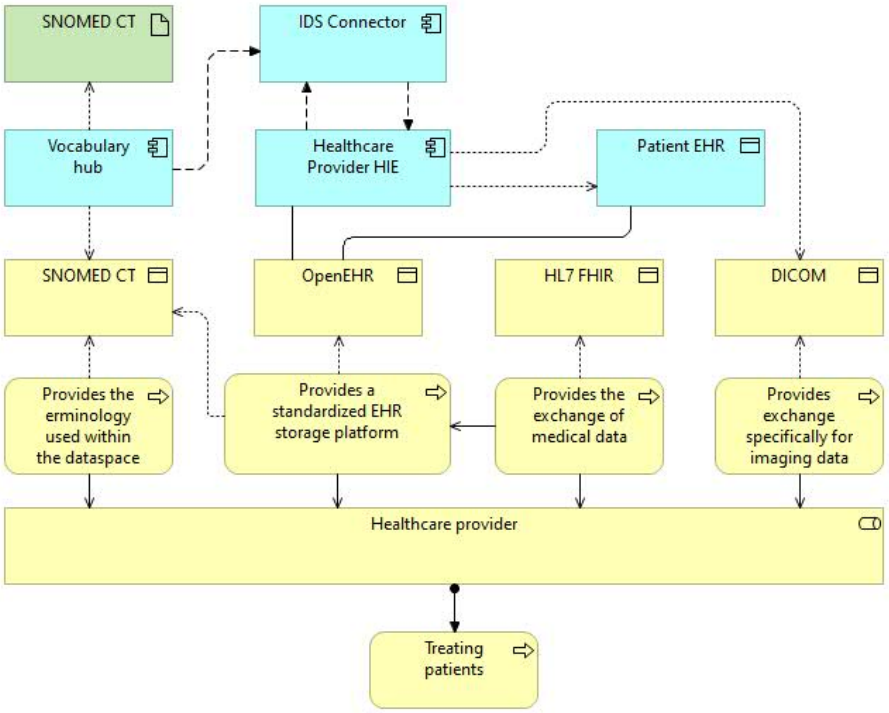
FIGURE 7.1: Annotated Dataspace process viewpoint



Updated Standards viewpoint:

Based on additional feedback, some changes were required of the standards model presented in chapter 4 (figure 4.13). This model is presented below:

FIGURE 7.2: Updated standards viewpoint



Compared to figure 4.13 this viewpoint includes more links to the dataspace by adding the HIE and IDS connector elements. The OpenEHR is linked to the HIE as the EHRs presented in the HIE are structured according to OpenEHR standards. DICOM is linked to the HIE and not the Patient EHR as the images are not directly stored in a patient EHR but linked to a patient using metadata. The images are viewed in the HIE using a HTML viewer of a PAC (Picture archiving and communication) system. This system is not modeled to keep the focus on the standards and the dataspace.

7.2.2 Applying in practice

Implementing the proposed dataspace in practice could be a great potential next step following this thesis. Doing so could validate its impact on interoperability, and provide numerical, quantitative evidence on efficiency and scalability. This section discusses what applying the architecture in practice could look like. In chapter 4 the last step of the ADM cycle which was included was step 5: Technology Architecture. When applying the thesis in practice, the following steps of the ADM cycle should be followed. However, in order to achieve this certain other elements need to be executed to ensure that the ADM based implementation goes well. This is done in a few steps:

1. **Practical testing of individual components:**
   Given that the architecture features several components which are yet to be developed, this need to be done first. Following that they need to be tested in a testing environment. The connector need to be able to communicate one on one first as a demo. Then the the individual components are added sequentially to ensure everything works. Special attention needs to be put on the communication between

existing healthcare systems and the connectors (step 3 in figure 7.1) since if this fails, the communication will fall apart instantly. Then when it is confirmed all components work as they are supposed to in a small one to one demo, then this can be expanded and tested further.

2. **Performance and scalability testing:**
   When everything is confirmed to work as it should, then performance and scalability tests need to be performed to ensure that the system can handle the loads that are required of it when in operation. This is also when the first quantitative tests can be done.

3. **Ensuring legislative compliance:**
   As medical data is highly sensitive an analysis needs to be done by an specialist in IT legislation. This is done to avoid the dataspace breaking any regulation which could bring large fines to the operators of the system. The performance of this step also fulfills requirement 8 in table 3.4.

4. **Deployment in Twente:**
   When the steps above are completed, which ensure that the system is ready for operation, then the next steps of the ADM cycle can be executed. These are Opportunities and Solutions, which is where a roadmap of the implementation is created (among other things). Then step 7, Migration Planning, during which a very detailed migration plan for the dataspace is created. And then finally, during step 8 called Implementation Governance, the migration is performed according to the migration plan created in step 7.

5. **Architecture Change Management:**
   The last step of the ADM cycle is Architecture Change Management. During this step you handle any problems that come up, and monitor whether the architecture fulfills the requirements and goals of the project.

This would complete the ADM cycle putting the dataspace in operation. The ADM also allows the architect to go back to another phase of the ADM to make any required changes when they come up.

7.2.3 Direction of future research in general

In general, the research regarding (medical) dataspaces is lacking. This means that any research on the topic is valuable. That being said, certain aspects could accelerate practical applications of medical dataspace technology. In 2022 the EU proposed the European Health Dataspace (EHDS). While the name includes the term dataspace, no architecture was suggested in this proposal. Future research could analyze the proposal, see how this aligns with current dataspace literature and identify gaps in the knowledge currently available. When these are identified, research can be executed specifically for filling this gap, making the path to realising a true EHDS more viable. So I suggest that the general direction of the field should focus on paving the way for the EHDS implementation. For this, one could think of researching the potential of decentralized dataspaces, where multiple smaller dataspaces are interoperable with each other (for example one per EU country in the EHDS).

That being said, the literature on the medical dataspace specifically is simply so minimal that any publication on the topic will proof to be novel for the foreseeable future.

7.3 Overall Conclusion

In this section a summary will be provided of the findings of this thesis, this will be done using various aspects which were discussed throughout.

- **Current situation:**
  Background literature research and stakeholder interviews revealed that in the current situation interoperability between healthcare organisations is not always appropriately facilitated in the Netherlands. Common healthcare processes, like GP to Hospital have their own specialized system and in general function well. However, there are many more organisations in Twente which produce medical data, which has to be exchanged on a regular basis. One of the interviewed stakeholder when asked about the current interoperability said "We fax [documents]". This indicates the current state of interoperability between healthcare providers. Secondly, researchers often still have to go through patient records to find the medical information to require for their research. An alternative approach for this process should be presented too.

- **Problem Investigation:**
  The problem investigation stage uncovered three problems within the scope of this research. These three problems are:

  - Lack of regional medical data exchange
  - Need for interoperability and data sovereignty
  - Patient privacy and regulatory compliance

  Problem verification using stakeholders from within Twente verified that the first two points certainly had room for improvement. The third presented an aspect of their work which had to be taken into consideration, but was not necessarily a limiting factor or problem.

- **Architecture:**
  This thesis presents an architecture for a dataspace to facilitate interoperability of healthcare providers in the Twente region of the Netherlands. This architecture is presented using ArchiMate models of the Business, Application and Technology layer. A variety of viewpoints were discussed in order to show a range of functionalities of the dataspace. The conclusion following the presentation of the architecture was that a medical dataspace architecture could be developed. The architecture fulfilled 62.1% of its requirements fully, 24.1% partially and 13.8% not at all. The thesis was not put into practice, future research could execute this at least on a prototype level however.

- **Validation:**
  The validation of the architecture was performed through five expert interviews. The five experts came from various backgrounds, but all had a high level of expertise regarding data architecture. Generally speaking the experts gave positive feedback about the architecture design. Especially the interoperability which can be achieved using the design got positive remarks. Also scalability got favorable feedback, with some points of criticism however. The largest points of improvement which were concluded from the expert interviews was the lack of focus on security elements in the models, and a missing check on data quality before the data is exchanged in the system.

An answer to the sub-questions formulated in section 1.4 can be found in their respective chapter. Here the main research question of this thesis is answered based on the information above.

Although the main research question is already partly answered in question 2.2 above, this conclusion can now add the information gained during expert validation and thus, answer the question in its entirety. To remind the reader, the main research question of this thesis is as follows:

How to design an architecture for a regional medical datahub which ensures interoperability between existing standards, follows (data-)regulations and allows for appropriate data governance policies regarding access, quality and compliance using the IDS-RAM?

This thesis has presented an architecture which expert validation has shown, ensures interoperability between standards and healthcare providers. It follows (data-)regulations and allows for data governance policies to be enacted regarding access, quality and compliance using the IDS-RAM. That being said, there are still improvements which can be made to the architecture, especially regarding adding and visualizing security elements and data quality checking. Furthermore, given that the architecture follows the principles put forth by the IDS-RAM it also allows for the patient to have full data sovereignty, meaning they have control over what happens to their own medical data. Additionally, the dataspace offers a way for researchers to access medical data for their research purposes. However, expert validation pointed out that this secondary medical data usage can be more pronounced in the architecture. All in all, the architecture promotes interoperability between healthcare organisations, creating the potential for a more pleasant healthcare process for both patients and healthcare professionals.

Bibliography

[1] K. Hardy, "Paleomedicine and the evolutionary context of medicinal plant use," *Revista Brasileira de Farmacognosia*, vol. 31, no. 1, pp. 1–15, Oct. 2020. DOI: 10.1007/s43450-020-00107-4. [Online]. Available: https://doi.org/10.1007/s43450-020-00107-4.

[2] H. Jin, Y. Luo, P. Li, and J. Mathew, "A review of Secure and Privacy-Preserving Medical Data Sharing," *IEEE Access*, vol. 7, pp. 61 656–61 669, Jan. 2019. DOI: 10.1109/access.2019.2916503. [Online]. Available: https://doi.org/10.1109/access.2019.2916503.

[3] A. Staff, *Taking the pulse of data and technology in modern healthcare*, Sep. 2023. [Online]. Available: https://arcadia.io/resources/taking-the-pulse-of-data-and-technology-in-modern-healthcare#:~:text=Hospitals%20produce%20an%20average%20of,for%20healthcare%20will%20reach%2036%25..

[4] *Health data: A holistic approach to unlock the value of health data.* [Online]. Available: https://www2.deloitte.com/be/en/pages/life-sciences-and-healthcare/articles/health-data.html.

[5] E. F. Codd, "A relational model of data for large shared data banks," *Communications of the ACM*, vol. 13, no. 6, pp. 377–387, Jun. 1970. DOI: 10.1145/362384.362685. [Online]. Available: https://doi.org/10.1145/362384.362685.

[6] E. Hechler, M. Weihrauch, and Y. Wu, *Terminology: data fabric and data mesh.* Jan. 2023, pp. 17–42. DOI: 10.1007/978-1-4842-9253-2\{_}2. [Online]. Available: https://doi.org/10.1007/978-1-4842-9253-2_2.

[7] M. J. Franklin, A. Halevy, and D. Maier, "From databases to dataspaces," *Sigmod Record*, vol. 34, no. 4, pp. 27–33, Dec. 2005. DOI: 10.1145/1107499.1107502. [Online]. Available: https://doi.org/10.1145/1107499.1107502.

[8] *Scopus preview - Scopus - Welcome to Scopus.* [Online]. Available: https://www.scopus.com/.

[9] S. Coughlin, D. Roberts, K. O'Neill, and P. Brooks, "Looking to tomorrow's healthcare today: a participatory health perspective," *Internal medicine journal*, vol. 48, no. 1, pp. 92–96, Jan. 2018. DOI: 10.1111/imj.13661. [Online]. Available: https://doi.org/10.1111/imj.13661.

[10] M. Cuggia and S. Combes, "The French Health Data Hub and the German Medical Informatics Initiatives: two national projects to promote data sharing in healthcare," *Yearbook of Medical Informatics*, vol. 28, no. 01, pp. 195–202, Aug. 2019. DOI: 10.1055/s-0039-1677917. [Online]. Available: https://doi.org/10.1055/s-0039-1677917.

[11] I. D. S. Association, *IDSA Data Spaces Radar*, Mar. 2024. [Online]. Available: https://www.dataspaces-radar.org/radar/.

[12] B. Otto, M. T. Hompel, and S. Wrobel, *Designing data spaces*. Jan. 2022. DOI: 10.1007/978-3-030-93975-5. [Online]. Available: https://doi.org/10.1007/978-3-030-93975-5.

[13] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific data*, vol. 3, no. 1, Mar. 2016. DOI: 10.1038/sdata.2016.18. [Online]. Available: https://doi.org/10.1038/sdata.2016.18.

[14] T. O. Group, *The ArchiMate® Enterprise Architecture Modeling Language*, 2024. [Online]. Available: https://www.opengroup.org/archimate-forum/archimate-overview.

[15] R. Wieringa, *Design Science Methodology for information systems and software engineering*. Jan. 2014. DOI: 10.1007/978-3-662-43839-8. [Online]. Available: https://doi.org/10.1007/978-3-662-43839-8.

[16] X. L. Dong and A. Halevy, "Indexing dataspaces," *SIGMOD'07*, Jun. 2007. DOI: 10.1145/1247480.1247487. [Online]. Available: https://doi.org/10.1145/1247480.1247487.

[17] B. Otto, *The evolution of data spaces*. Jan. 2022, pp. 3–15. DOI: 10.1007/978-3-030-93975-5\{_}1. [Online]. Available: https://doi.org/10.1007/978-3-030-93975-5_1.

[18] "Aligning Dutch Logistics Data Spaces initiatives to the international Data Spaces: Discussing the state of development," Valencia, Spain, Mar. 2022.

[19] I. D. S. Association, *Why - International Data spaces*, Mar. 2024. [Online]. Available: https://internationaldataspaces.org/why/.

[20] I. D. Association, "IDS-RAM 4," Tech. Rep. 4.0, 2022.

[21] F. Amato, *Luxembourg launches Dataspace 4 Health 187; Luxembourg Institute of Health*, May 2024. [Online]. Available: https://www.lih.lu/en/article/luxembourg-launches-dataspace-4-health-a-pioneering-dataspace-and-governance-framework-for-secure-and-compliant-health-data-exchange/.

[22] G.-X. E. A. for Data and C. AISBL, *Home - Gaia-X: A federated Secure data infrastructure*, Feb. 2024. [Online]. Available: https://gaia-x.eu/.

[23] Health-X, *HEALTH-X dataLOFT Platform*, 2024. [Online]. Available: https://www.health-x.org/en/platform.

[24] Health-X, *HEALTH-X dataLOFT Platform - Milestones*, 2024. [Online]. Available: https://www.health-x.org/en/milestones.

[25] G. PROJECT, *About Gatekeeper | GATEKEEPER PROJECT*, Oct. 2023. [Online]. Available: https://www.gatekeeper-project.eu/about-gatekeeper/#block-missionandvision.

[26] Fraunhofer, *Fraunhofer Medical Data Space*, 2023. [Online]. Available: https://www.medical-data-space.fraunhofer.de/en.html.

[27] F. Vijselaar, P. Jeekel, W. Niessen, R. Roozendaal, and N. van Meeteren, "Artificial intelligence in health," Tech. Rep., 2021. [Online]. Available: https://www.rvo.nl/sites/default/files/2022/03/AI%20LSH.pdf.

[28] C. Cordis, *VELES Excellence Hub - Strengthening the South-East Europe Smart Health Regional Excellence and Boosting the Innovation Potential*, Jul. 2023. [Online]. Available: https://cordis.europa.eu/project/id/101087483.

[29] T. Benson and G. Grieve, *Principles of health interoperability*, 4th ed. Springer, Jan. 2021. DOI: 10.1007/978-3-030-56883-2. [Online]. Available: https://doi.org/10.1007/978-3-030-56883-2.

[30] H. International, "2024 State of FHIR survey results," Tech. Rep., May 2024. [Online]. Available: https://www.hl7.org/documentcenter/public/white-papers/2024%20StateofFHIRSurveyResults_final.pdf.

[31] DICOM, "DICOM PS3.1 2024B - Introduction and Overview," Tech. Rep. DICOM PS3.1 2024b, 2024. [Online]. Available: https://dicom.nema.org/medical/dicom/current/output/pdf/part01.pdf.

[32] E. R. Ranschaert and F. H. B. Binkhuysen, "European Teleradiology now and in the future: results of an online survey," *Insights into Imaging*, vol. 4, no. 1, pp. 93–102, Dec. 2012. DOI: 10.1007/s13244-012-0210-z. [Online]. Available: https://doi.org/10.1007/s13244-012-0210-z.

[33] Nictiz, *SNOMED CT - Nictiz*, Sep. 2023. [Online]. Available: https://nictiz.nl/standaarden/overzicht-van-standaarden/snomed-ct/.

[34] OpenEHR, *OpenEHR - About US*, 2024. [Online]. Available: https://www.openehr.org/about_us.

[35] K. Atalag, T. Beale, R. Chen, T. Gornik, S. Heard, and I. McNicoll, "openEHR – A Semantically-enabled Health Computing Platform," Tech. Rep. [Online]. Available: https://www.openehr.org/static/files/resources/openEHR_vendor_independent_platform.pdf.

[36] *Home | AORTA-LSP*, 2024. [Online]. Available: https://www.aorta-lsp.nl/.

[37] C. voor Ethiek en Gezondheid, *Argumentenwijzer over Elektronische Patiënten Dossiers*, pp. 2–3. [Online]. Available: https://www.ceg.nl/binaries/ceg/documenten/publicaties/2013/04/04/argumentenwijzer-over-elektronische-patienten-dossiers/Argumentenwijzer20EPD.pdf.

[38] W. e. S. Ministerie van Volksgezondheid, *About CEG*, Mar. 2022. [Online]. Available: https://www.ceg.nl/about-ceg.

[39] CumuluZ, *CumuluZ | Eén data-infrastructuur. Ván de zorg, vóór de zorg.* 2024. [Online]. Available: https://www.cumuluz.org/.

[40] E. MC, *Digizorg app voor patiënten.* [Online]. Available: https://www.erasmusmc.nl/nl-nl/digizorg-app.

[41] E. MC, *Veelgestelde vragen over de Digizorg patiënten app*, 2024. [Online]. Available: https://www.erasmusmc.nl/nl-nl/digizorg-app/veelgestelde-vragen#83d61e9d-d3be-4713-9979-a183d5ccc391.

[42] R. Noord, *Zorgviewer | RIVO Noord*, Mar. 2024. [Online]. Available: https://www.rivo-noord.nl/zorgviewer/.

[43] *Healthdataspaceamsterdam.nl*, Mar. 2024. [Online]. Available: https://www.healthdataspaceamsterdam.nl/.

[44] Health-RI, *About Health-RI | Health-RI*, 2024. [Online]. Available: https://www.health-ri.nl/en/about-health-ri.

[45] Health-RI, *Services tools | Health-RI*, 2024. [Online]. Available: https://www.health-ri.nl/en/services.

[46] *Over Nictiz - Nictiz*, Jun. 2024. [Online]. Available: https://nictiz.nl/over-nictiz/.

[47] *ZIRA | Ziekenhuis Referentie architectuur - NICTIZ*, Oct. 2023. [Online]. Available: https://nictiz.nl/standaarden/referentiedomeinenmodellen/zira/.

[48] L. M. Bollweg, *Data governance for managers*. Jan. 2022. DOI: 10.1007/978-3-662-65171-1. [Online]. Available: https://doi.org/10.1007/978-3-662-65171-1.

[49] I. D. S. Association, *3.1.1 Roles in the International Data Spaces | IDS Knowledge Base*, Feb. 2023. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/ids-ram-4/layers-of-the-reference-architecture-model/3-layers-of-the-reference-architecture-model/3-1-business-layer/3_1_1_roles_in_the_ids.

[50] M. L. Zeng and J. Qin, *Metadata, Second edition*, 2nd ed. Feb. 2016. [Online]. Available: http://digitalcommons.kent.edu/facultybooks/64/.

[51] P. Buneman, S. Khanna, and T. Wang-Chiew, "Why and where: A characterization of data provenance," in *Database Theory — ICDT 2001*, J. Van den Bussche and V. Vianu, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 316–330.

[52] A. D. Leander, "Identification and implementation of suitable decentralized identifier methods for self-sovereign identity wallets in a data space," Master's Thesis, University of Twente, 2024. [Online]. Available: https://essay.utwente.nl/103725/.

[53] I. D. S. Association, "IDS-RAM 4: Identity and Trust Management," Tech. Rep., Apr. 2022.

[54] E. Union, *Regulation - 2016/679 - EN - gdpr - EUR-Lex*, May 2016. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[55] *Data protection in the EU*, Jul. 2023. [Online]. Available: https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_en.

[56] E. Union, *Types of legislation | European Union*. [Online]. Available: https://european-union.europa.eu/institutions-law-budget/law/types-legislation_en.

[57] *Data protection under GDPR - Your Europe*, Jul. 2022. [Online]. Available: https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/index_en.htm#:~:text=The%20GDPR%20sets%20out%20detailed,people%20living%20in%20the%20EU..

[58] B. Wolford, *What are the GDPR fines?* Sep. 2023. [Online]. Available: https://gdpr.eu/fines/#:~:text=Under%20the%20GDPR%2C%20fines%20are,the%20severity%20of%20the%20penalty..

[59] *Fines / penalties - General Data Protection Regulation (GDPR)*, Oct. 2021. [Online]. Available: https://gdpr-info.eu/issues/fines-penalties/#:~:text=For%20especially%20severe%20violations%2C%20listed,fiscal%20year%2C%20whichever%20is%20higher..

[60] E. D. P. Board, *1.2 billion euro fine for Facebook as a result of EDPB binding decision | European Data Protection Board*, May 2023. [Online]. Available: https://www.edpb.europa.eu/news/news/2023/12-billion-euro-fine-facebook-result-edpb-binding-decision_en.

[61] W. e. S. Ministerie van Volksgezondheid, "Wat u moet weten over de Wegiz | Wet en Aanleiding," Tech. Rep., Nov. 2021. [Online]. Available: https://www.zorginzicht.nl/binaries/content/assets/zorginzicht/algemeen-ondersteuning/wat-u-moet-weten-over-de-wegiz---wet-en-aanleiding.pdf.

[62] W. e. S. Ministerie van Volksgezondheid, "Wat u moet weten over de Wegiz | Hoe de Wegiz werkt," Tech. Rep., Nov. 2021. [Online]. Available: https://www.zorginzicht.nl/binaries/content/assets/zorginzicht/algemeen-ondersteuning/wat-u-moet-weten-over-de-wegiz---hoe-de-wegiz-werkt.pdf.

[63] E. K. der Staten-Generaal, *Uitvoeringswet Algemene verordening gegevensbescherming (34.851)*, May 2018. [Online]. Available: https://www.eerstekamer.nl/wetsvoorstel/34851_uitvoeringswet_algemene.

[64] W. e. S. Ministerie van Volksgezondheid, *Wabvpz*, Jul. 2023. [Online]. Available: https://www.avghelpdeskzorg.nl/onderwerpen/wabvpz.

[65] N. N. Instituut, *NEN 7510: Informatiebeveiliging in de zorg - ICT in de zorg - Zorg Welzijn*, 2024. [Online]. Available: https://www.nen.nl/zorg-welzijn/ict-in-de-zorg/informatiebeveiliging-in-de-zorg.

[66] W. e. S. Ministerie van Volksgezondheid, "AVG-Helpdesk voor Zorg, Welzijn en Sport," Tech. Rep. [Online]. Available: https://www.avghelpdeskzorg.nl/binaries/avghelpdeskzorg/documenten/publicaties/2024/01/15/infographic-wet--en-regelgeving/VWS_AVGHelpdesk_interactief_DEF.pdf.

[67] N. N. Instituut, *NEN 7513:2023*, Jul. 2023. [Online]. Available: https://www.nen.nl/artnr/309856.

[68] RIVM, *Wet geneeskundige behandelingsovereenkomst (WGBO)*, Dec. 2021. [Online]. Available: https://www.rivm.nl/cpt/kwaliteit-wet-en-regelgeving/wetgeving/wgbo.

[69] *wetten.nl - Regeling - Wet op de beroepen in de individuele gezondheidszorg - BWBR0006251*, Jan. 2024. [Online]. Available: https://wetten.overheid.nl/BWBR0006251/2024-01-01.

[70] Mitz, *Mitz, de online toestemmingsvoorziening*, 2024. [Online]. Available: https://www.mitz-toestemming.nl/.

[71] Mitz, *Eerste zorgaanbieder aangesloten op Mitz, de online toestemmingsvoorziening | Mitz Toestemming*, Nov. 2023. [Online]. Available: https://www.mitz-toestemming.nl/actueel/nieuws/eerste-zorgaanbieder-aangesloten-op-mitz-de-online-toestemmingsvoorziening.

[72] R. Bild, K. A. Kuhn, and F. Prasser, "Better Safe than Sorry - Implementing Reliable Health Data Anonymization.," *PubMed*, vol. 270, pp. 68–72, Jun. 2020. DOI: 10.3233/shti200124. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/32570348.

[73] O. Vovk, G. Piho, and P. Ross, "Methods and tools for healthcare data anonymization: a literature review," *International Journal of General Systems*, vol. 52, no. 3, pp. 326–342, Feb. 2023. DOI: 10.1080/03081079.2023.2173749. [Online]. Available: https://doi.org/10.1080/03081079.2023.2173749.

[74] F. Bettahar, C. Moulin, and J.-P. Barthes, "Towards a Semantic Interoperability in an e-Government Application," *Electronic Journal of e-Government*, vol. 7, no. 3, Nov. 2009. [Online]. Available: https://academic-publishing.org/index.php/ejeg/article/view/503.

[75] T. W. Cook, J. R. d. M. Nogueira, and L. T. Cavalini, *Knowledge management of controlled vocabularies in healthcare and the semantic interoperability challenge*. 2015, pp. 57–78, Cited by: 0. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84956777761&partnerID=40&md5=e85146d0f9b69b0a6e8a3aca477

[76] P. White and A. Roudsari, "An ontology for healthcare quality indicators: Challenges for semantic interoperability," *Studies in Health Technology and Informatics*, vol. 210, pp. 414–418, 2015, Cited by: 3. DOI: 10.3233/978-1-61499-512-8-414. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84937459776&doi=10.3233%2f978-1-61499-512-8-414&partnerID=40&md5=0a5ae436f6f3c8ad0e7764954b7fe4fb.

[77] C. Martínez-Costa, M. Menárguez-Tortosa, and J. T. Fernández-Breis, "An approach for the semantic interoperability of iso en 13606 and openehr archetypes," *Journal of Biomedical Informatics*, vol. 43, no. 5, pp. 736–746, 2010, ISSN: 1532-0464. DOI: https://doi.org/10.1016/j.jbi.2010.05.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1532046410000821.

[78] B. H. de Mello, S. J. Rigo, C. A. da Costa, *et al.*, "Semantic interoperability in health records standards: A systematic literature review," *Health and Technology*, vol. 12, no. 2, pp. 255–272, 2022, Cited by: 55; All Open Access, Bronze Open Access. DOI: 10.1007/s12553-022-00639-w. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85123640532&doi=10.1007%2fs12553-022-00639-w&partnerID=40&md5=eaeafe45407b69f79bcc056ee2689736.

[79] H. van der Veer and A. Wiles, "Achieving Technical Interoperability - the ETSI Approach," Tech. Rep. 3, Apr. 2008. [Online]. Available: https://www.etsi.org/images/files/ETSIWhitePapers/IOP%20whitepaper%20Edition%203%20final.pdf.

[80] *GitHub: Let's build from here*, 2024. [Online]. Available: https://github.com/.

[81] A. Jacobsen, R. de Miranda Azevedo, N. Juty, *et al.*, "FAIR Principles: Interpretations and Implementation Considerations," *Data Intelligence*, vol. 2, no. 1-2, pp. 10–29, Jan. 2020, ISSN: 2641-435X. DOI: 10.1162/dint_r_00024. eprint: https://direct.mit.edu/dint/article-pdf/2/1-2/10/1893430/dint_r_00024.pdf. [Online]. Available: https://doi.org/10.1162/dint%5C_r%5C_00024.

[82] *TOGAF® usage worldwide*. [Online]. Available: https://www.opengroup.org/togaf-usage-worldwide.

[83] R. Alm and M. Wißotzki, *TOGAF adaption for small and medium enterprises*. Jan. 2013, pp. 112–123. DOI: 10.1007/978-3-642-41687-3\{_}12. [Online]. Available: https://doi.org/10.1007/978-3-642-41687-3_12.

[84] A. E. A. Sanyoto and M. C. Saputra, "ArchiMate's Strengths and Weaknesses as EA Modeling Language: A Systematic Mapping Study," Dec. 2023. DOI: 10.1109/icic60109.2023.10381985. [Online]. Available: https://doi.org/10.1109/icic60109.2023.10381985.

[85] I. F. Alexander, "A taxonomy of stakeholders," *International journal of technology and human interaction*, vol. 1, no. 1, pp. 23–59, Jan. 2005. DOI: 10.4018/jthi.2005010102. [Online]. Available: https://doi.org/10.4018/jthi.2005010102.

[86] M. Jobst and T. Fischer, *Approaching a Common Conscious Dataspace from a Data Provider Perspective – Requirements and Perspectives*. Jan. 2022, pp. 333–343. DOI: 10.1007/978-3-031-10450-3\{_}28. [Online]. Available: https://doi.org/10.1007/978-3-031-10450-3_28.

[87] T. Dam, A. Krimbacher, and S. Neumaier, "Policy patterns for usage control in data spaces," Tech. Rep., Sep. 2023. [Online]. Available: https://arxiv.org/pdf/2309.11289.

[88] D. R. Firdausy, P. De Alencar Silva, M. Van Sinderen, and M.-E. Iacob, "Towards a Reference Enterprise Architecture to enforce Digital Sovereignty in International Data Spaces," Jun. 2022. DOI: 10.1109/cbi54897.2022.00020. [Online]. Available: https://doi.org/10.1109/cbi54897.2022.00020.

[89] K. Talmoudi, K. Choukri, and I. Gavanon, "Compliance by Design Methodologies in the Legal Governance Schemes of European Data Spaces," Tech. Rep. 2024.legal-1.1, May 2024. [Online]. Available: https://aclanthology.org/2024.legal-1.1.

[90] C. Luidold and C. Jungbauer, "Cybersecurity policy framework requirements for the establishment of highly interoperable and interconnected health data spaces," *Frontiers in medicine*, vol. 11, May 2024. DOI: 10.3389/fmed.2024.1379852. [Online]. Available: https://doi.org/10.3389/fmed.2024.1379852.

[91] D. R. Firdausy, P. De Alencar Silva, M. Van Sinderen, and M.-E. Iacob, *A data connector store for international data spaces*. Jan. 2022, pp. 242–258. DOI: 10.1007/978-3-031-17834-4\{_}14. [Online]. Available: https://doi.org/10.1007/978-3-031-17834-4_14.

[92] P. Hagenhoff, S. Biehs, F. Möller, and B. Otto, *Designing a reference architecture for collaborative condition monitoring data spaces: design requirements and views*. Jan. 2024, pp. 355–369. DOI: 10.1007/978-3-031-61175-9\{_}24. [Online]. Available: https://doi.org/10.1007/978-3-031-61175-9_24.

[93] R. A. Deshmukh, S. A. Chala, and C. Lange, "Requirements and Building Blocks for Manufacturing Dataspaces," Apr. 2023. DOI: 10.1145/3543873.3587664. [Online]. Available: https://doi.org/10.1145/3543873.3587664.

[94] I. D. S. Association, *Rule Book | IDS Knowledge Base*, 2023. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/1_introduction.

[95] I. D. S. Association, *Functional Requirements | IDS Knowledge Base*, Apr. 2024. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/idsa-rulebook/idsa-rulebook/3_functional_requirements.

[96] T. O. Group, *The TOGAF® Standard, Version 9.2*, Apr. 2018. [Online]. Available: https://pubs.opengroup.org/architecture/togaf9-doc/arch/.

[97] W. e. S. Ministerie van Volksgezondheid, "Integraal Zorgakkoord," Tech. Rep., Sep. 2022. [Online]. Available: https://www.dejuistezorgopdejuisteplek.nl/.uc/f088c384b0102bb01b60076df250226d850bb057b90e700/integraal-zorg-akkoord.pdf.

[98] M. van VWS, *Integraal Zorgakkoord Samen werken aan gezonde zorg - Infographic*, Sep. 2022. [Online]. Available: https://open.overheid.nl/documenten/ronl-d6cdb51e2b6363daf11f82b5ae083dbee263692b/pdf.

[99] M. Bloch, S. Blumberg, J. Laartz, and M. Online, "Delivering large-scale IT projects on time, on budget, and on value," Tech. Rep., Oct. 2012. [Online]. Available: https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/delivering-large-scale-it-projects-on-time-on-budget-and-on-value#/.

[100] F. F. Silveira, R. De F S Macri Russo, I. G. Júnior, and R. Sbragia, *Systematic review of risks in domestic and global IT projects*. Dec. 2021, pp. 1612–1634. DOI: 10.4018/978-1-6684-3702-5.ch076. [Online]. Available: https://doi.org/10.4018/978-1-6684-3702-5.ch076.

[101] D. Firdausy, "Designing essential components for logistics data spaces: Connecting logistics interfaces, converters, knowledge, and standards," Tech. Rep., Jul. 2023. DOI: 10.3990/1.9789036557177. [Online]. Available: https://doi.org/10.3990/1.9789036557177.

[102] I. D. S. Association and Fraunhofer, "DIN SPEC 27070:2020-03," Tech. Rep. 27070:2020-03, Mar. 2020. [Online]. Available: https://internationaldataspaces.org/ids-is-officially-a-standard-din-spec-27070-is-published/#:~:text=DIN%20SPEC%2027070%20specifies%20the,architecture%20and%20cyber%20security%20measures..

[103] J. Piest, S. Slavova, and W. van Heeswijk, "A reference use case, data space architecture, and prototype for smart truck parking," English, in *Proceedings of the 22nd CIAO! Doctoral Consortium, and Enterprise Engineering Working Conference Forum 2022 co-located with 12th Enterprise Engineering Working Conference (EEWC 2022)*, ser. CEUR Workshop Proceedings, 22nd Doctoral Consortium, and Enterprise Engineering Working Conference Forum, CIAO 2022, CIAO 2022 ; Conference date: 02-11-2022 Through 03-11-2022, CEUR, May 2023, pp. 1–15.

[104] TwenteBeter, *Regiovisie: TwenteBeter*. [Online]. Available: https://www.twentebeter.nl/.

[105] R. Twente, "Samen koersen op gezondheid en goede zorg in Twente," Tech. Rep., Dec. 2023. [Online]. Available: https://www.twentebeter.nl/regioplan.pdf.

[106] O. Beyan, A. Choudhury, J. van Soest, *et al.*, "Distributed analytics on sensitive medical data: The personal health train," *Data Intelligence*, vol. 2, no. 1-2, pp. 96–107, 2020. DOI: 10.1162/dint_a_00032. [Online]. Available: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85117792651&doi=10.1162%2fdint_a_00032&partnerID=40&md5=437d9d7fa7ed132ad32ef6e8abd32344.

Appendix A

Appendix A - List of ICT standards in use in the Netherlands

Standard: *GMDN*
Description: A system and standard which is used for the naming and categorisation of all kinds of medical devices and products. It uses a database which lists all terms and also provides codes and definitions for these terms.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://www.gmdnagency.org/what-we-do/

Standard: *LOINC*
Description: A common language mostly focussed on identifying a variety of health measurements and observations. It allows for both sender and receiver of the data to have a common understanding of what the observations and measurements mean.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://loinc.org/about/

Standard: *SNOMED CT*
Description: A comprehensive system of healthcare terminology. Each term has a concept code, description and can have relationships to another SNOMED CT term (e.g. Infective Pneumonia IS A Respiratory Disease). It provides the core general terminology for EHRs.
In context of medical hub: Very common communication standard, used by many organisations in EHRs (e.g. at the NHS), thus can be used very well in a medical data hub for communication between various parties.
Reference: https://www.snomed.org/files/ugd/9002748a849a3565054d14a4c94cf1062331a3

Standard: *ATC*
Description: Is a standard for medicine maintained by the World Health Organisation (WHO). It classifies each active ingredient by which organ it affects. The codes can be utilized by IT systems to communicate specific medicines or when doing research to ensure that the correct medicine is being examined. Example of a code is A10BA02 (metformin).
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://www.who.int/standards/classifications/other-classifications/the-anatomical-therapeutic-chemical-classification-system-with-defined-daily-doses

Standard: *DBC*

Description: A Dutch standard. It provides a code for a complete treatment process. As the code encapsulates the entire diagnosis to treatment process it is used to determine what the costs are to the patient for their hospital visit.
In context of medical hub: Limited relevance. Potentially interesting for patients to know how the bill to their insurer is structured
Reference: https://www.zorgwijzer.nl/faq/dbc

Standard: *DSM*
Description: Is a handbook for mental health professionals for the diagnosis of various mental health disorders. Amongst other things it provides a common language. This common language ensures that both mental healthcare professionals and researchers are using the same semantics.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://www.psychiatry.org/psychiatrists/practice/dsm/frequently-asked-questions

Standard: *G-Standaard*
Description: A Dutch standard. It is a database which contains all the products which are provided by a pharmacy.
In context of medical hub: Limited relevance
Reference: https://www.knmp.nl/over-de-knmp/producten-en-diensten/wat-is-de-g-standaard

Standard: *ICD*
Description: An international standard for classification of diseases. It is maintained by the WHO and provides codes for various diseases and their symptoms. It is designed to map certain conditions to categories.
In context of medical hub: Limited relevance, could be used to communicate information in EHRs
Reference: https://www.who.int/standards/classifications/classification-of-diseases

Standard: *ICF*
Description: Is an WHO international standard concerning disability and functioning of an individual.
In context of medical hub: Limited relevance, could potentially be used to communicate information in EHRs
Reference: https://www.who.int/standards/classifications/international-classification-of-functioning-disability-and-health

Standard: *KMEHR/SumEHR*
Description: A standard introduced to structure clinical data. The SumEHR is a summarzied version of a KMEHR message. This standard is popularized by the Belgian government.
In context of medical hub: Could used as a data exchange standard in the hub
Reference: https://www.ehealth.fgov.be/standards/kmehr/en

Standard: *ICNP*
Description: An international standard for the nursing practice. It provides terms for nurses to use for both observations and interventions on patients.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://www.icn.ch/how-we-do-it/projects/ehealth-icnptm/about-icnp

Standard: *ICPC*
Description: International standard maintained by the WHO. It is centered around primary care encounters it allows for practicioners to classify and code four key elements: reason for the encounter, the diagnosis, functioning and processes of care.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://www.knmp.nl/over-de-knmp/producten-en-diensten/wat-is-de-g-standaard

Standard: *NHG-Standaarden*
Description: A Dutch standard. It provides general practioners guidelines for diagnosis and treatment of an array of common symptoms in a general practioners office. Each standard foccuses on one specific problem, complaint or risk factor and provides the doctor with guidelines with how to handle the patient adequately
In context of medical hub: Limited relevance
Reference: https://richtlijnen.nhg.org/over-nhg-richtlijnen

Standard: *NIC*
Description: The NIC provides classfication of treatments which nurses may perform. It has standardized language for both the treatments which are initiated by the nurse or the doctor.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://pubmed.ncbi.nlm.nih.gov/8591448/

Standard: *NOC*
Description: A taxonomy which is designed to classify certain patient outcomes which have been influenced by nursing care. The NOC contains 330 outcomes, and each with a label, a definition, and a set of indicators and measures to determine achievement of the nursing outcome.
In context of medical hub: Limited relevance
Reference: https://pubmed.ncbi.nlm.nih.gov/9610010/

Standard: *HL7*
Description: Is a range of application level standards which aim to facilitate (health) data transfer between applications. It is very widely used and has various different versions.
In context of medical hub: HL7 is very widely adopted and has extensive documentation. Probably a good choice for health data communication standard in the hub.
Reference: https://www.hl7.org/about/index.cfm?ref=nav

Standard: *HL7 FHIR*
Description: Fast Healthcare Interoperability Resources - version of HL7. It is a HL7 standard based on RESTful operations and is also becoming increasingly more popular.
In context of medical hub: HL7 is very widely adopted and has extensive documentation. Probably a good choice for health data communication standard in the hub.
Reference: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=491

Standard: *Omaha system*
Description:A Dutch standard. Is a classification system mostly used in elderly care. It is a terminology and code system. It used to log the health, behavior and measurements of a client.

In context of medical hub: Could potentially be used to connect elderly care data to the regional hub
Reference: https://www.omahasystem.nl/over-omaha-system/werken-met-het-omaha-system

Standard: *GS1*
Description: Is an organisation focused on the design and implantation of standards for communication between organisations. Amongst other things they are responsible for barcodes.
In context of medical hub: Limited relevance in this context
Reference: https://www.gs1.org/about

Standard: *ISO 3166-1*
Description: Is a standard containing standardized codes for countries and their subdivisions around the world. It is maintained by ISO. Example US = United States, NL = The Netherlands
In context of medical hub: Given the regionallity of the hub, limited relevance
Reference: https://www.iso.org/iso-3166-country-codes.html

Standard: *AGB-code*
Description: A Dutch standard. It is a unique code which identifies different medical or healthcare organisations in the Netherlands. It contains 8 numbers, the first two identifying what type of organisation it is. Example: MST = 17082293
In context of medical hub: Can be used to identify organisations in the hub
Reference: https://www.vektis.nl/agb-register

Standard: *BIG-register*
Description: A Dutch standard. It is a register where everyone working in a large number of healthcare positions in the Netherlands has to be registered. Everyone gets his or her own number in this register. The BIG-register came forth from the "Wet BIG".
In context of medical hub: Can be used to identify personnel in the hub
Reference: https://www.bigregister.nl/over-het-big-register

Standard: *UZI-register*
Description: A Dutch standard. UZI makes it possible for healthcare providers to have their own unique identification. To do this they provide cards to healthcare organisations and their employees.
In context of medical hub: Limited relevance
Reference: https://www.uziregister.nl/

Standard: *DICOM*
Description: Is an international standard for transmitting, storing, retreiving, printing, processing and displaying medical images. One of the primary goals for the standard is to facilitate communication between various interested parties, both hardware and software related.
In context of medical hub: Very common standard for images and thus required to achieve full functionality of the hub
Reference: https://dicom.nema.org/medical/dicom/current/output/pdf/part01.pdf

Standard: *EDIFACT*

Description: Is the most widely used EDI document standard in Europe. The standard provides three main elements: syntax rules to structure the data which is being send, an exchange protocol and some standard messages.
In context of medical hub: Useful data exchange standard
Reference: https://edicomgroup.com/learning-center/edi/standards/edifact: :text=What%20is%20EDIFA

Standard: *DCM/ISO 13972*
Description: ISO 13972:2022, specifies clinical information models as concepts that can be used to structure information.
In context of medical hub: Can be used in EHRs to communicate information
Reference: https://www.iso.org/standard/79498.html

Standard: *EN13606/ISO13606-5:2019*
Description: Specifies a means for communicating EHR information.
In context of medical hub: The provided information structure can be used in the hub
Reference: https://www.iso.org/standard/67868.html

Standard: *LBZ*
Description: Basic registration of patient information that hospitals record for patients which have gone through a hospital stay or visit.
In context of medical hub: Can be used to ingest patient personal information
Reference: https://bronnen.zorggegevens.nl/Bron?naam=Landelijke-Basisregistratie-Ziekenhuiszorg

Standard: *openEHR*
Description: Standard specification for health data which you can store, retreive and exchange data using electronic health records.
In context of medical hub: Can be used for the EHRs in the hub
Reference: https://openehr.org/about_us

Standard: *PWD*
Description: A Dutch standard. The standard is targeted at facilitating communication regarding healthcare surrounding the birth process. The aim is to help the communication between caretaker and clients. It does this by having agreements in place on five different levels.
In context of medical hub: Limited relevance
Reference: https://www.zorginzicht.nl/kwaliteitsinstrumenten/pwd

Standard: *XDS*
Description: Intended to facilitate the sharing of medical documents and images among collaborating healthcare institutions, in a standardized and secure manner
In context of medical hub: Can be used to share images in the hub
Reference: https://nictiz.nl/standaarden/overzicht-van-standaarden/xds/

Standard: *AORTA*
Description: AORTA is the Dutch healthcare infrastructure which enables electronic sharing of patient data. This is used by hospitals and general practitioners to retrieve each others data. Key to AORTA is that the data is not stored centrally but on healthcare providers own systems. The data is then retrieved via LSP (see last row of this table).
In context of medical hub: The system (in collaboration with LSP) can be used for

the exchanging data between GPs and hospitals
Reference: https://www.aorta-lsp.nl/over-aorta-lsp

Standard: *LSP*
Description: LSP enables the sharing of patient data which is stored locally at care providers. Through the LSP care providers can request data from systems at other health-care providers.
In context of medical hub: The system (in collaboration with AORTA) can be used for the exchanging data between GPs and hospitals
Reference: https://www.aorta-lsp.nl/over-aorta-lsp

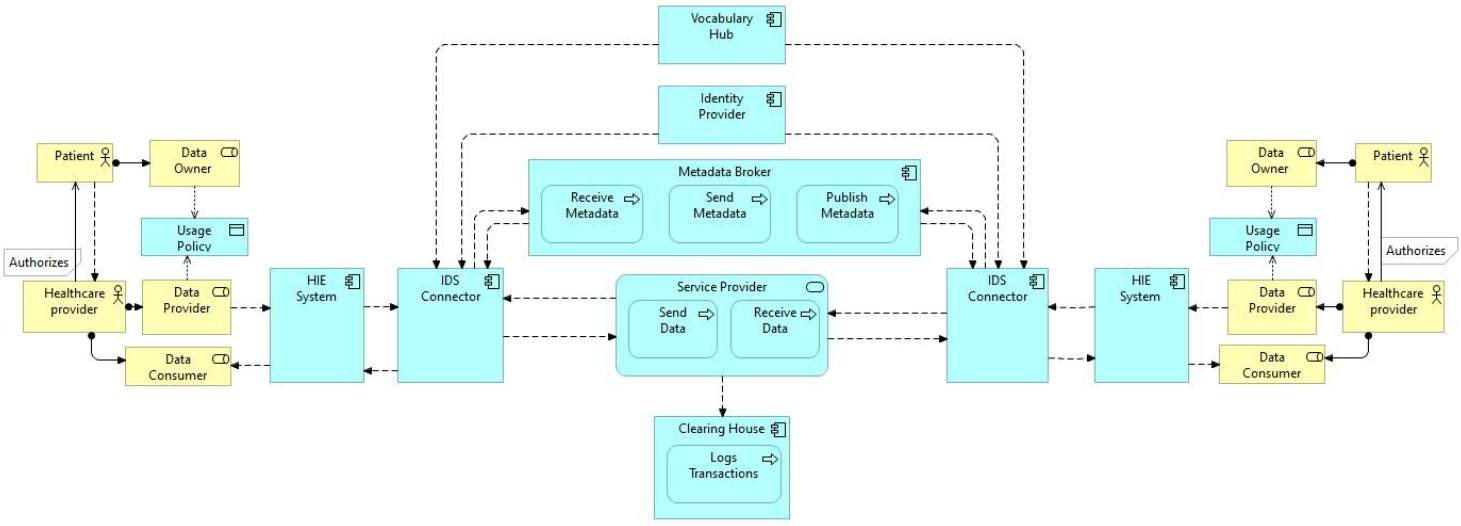
Appendix B

Appendix B - Template for expert validation interviews

Please note that only one of the interviews was held in English. The rest was done in Dutch using translated versions of the questions below.

- Thank you again for agreeing to doing this interview. First off, would you mind if I record this interview for reference later while processing the results? After processing is completed the recording will be deleted.
- Introduce the project, depending on familiarity of the interviewee

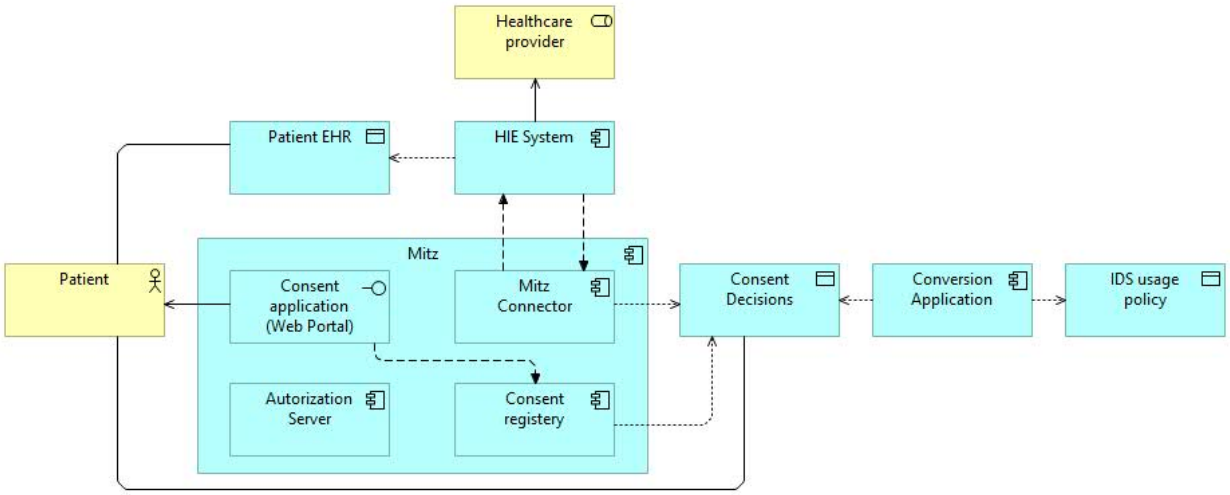
- Show dataspace model:



- This is a representation of the dataspace. It contains two representations of healthcare providers (an adjusted model was used for the interviews to make the model more clear on first look) with the their corresponding roles in the dataspace. Furthermore, it contains all the required elements to setup the dataspace. Then I go over all elements briefly.

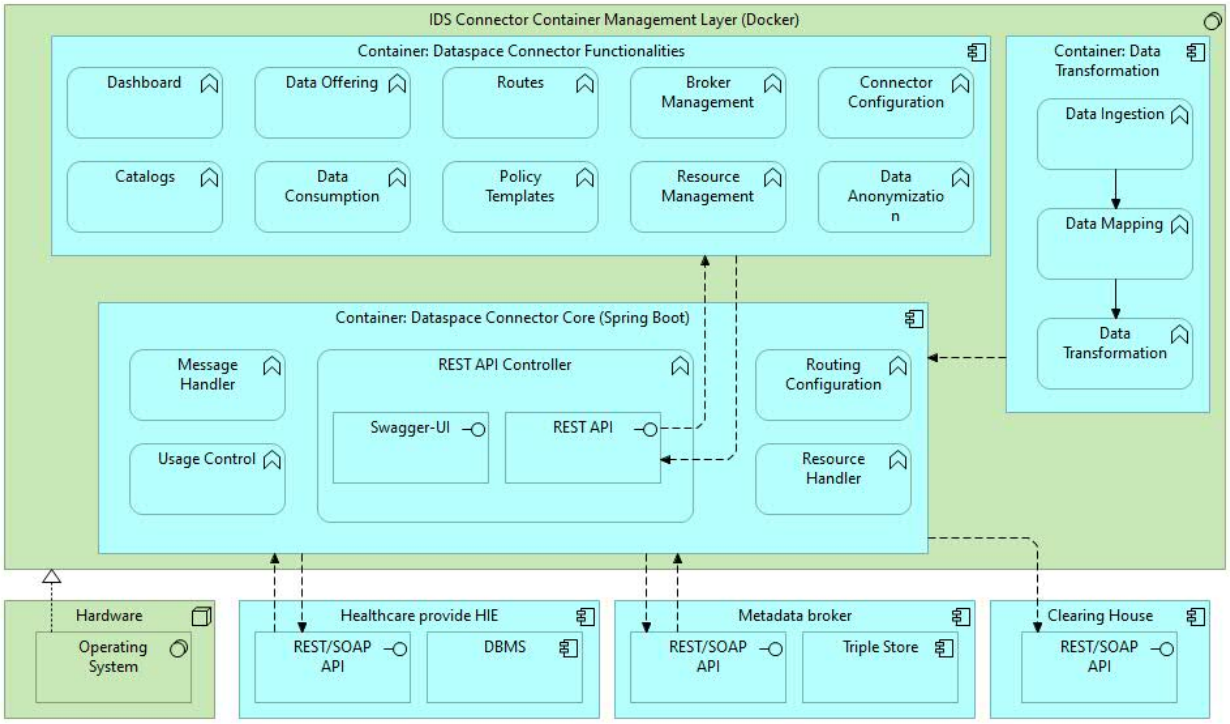
- Question: Does this model cover all the key aspects you would expect in a regional medical dataspace? If not, what would you say you are missing?

- Show Usage Policy creation model



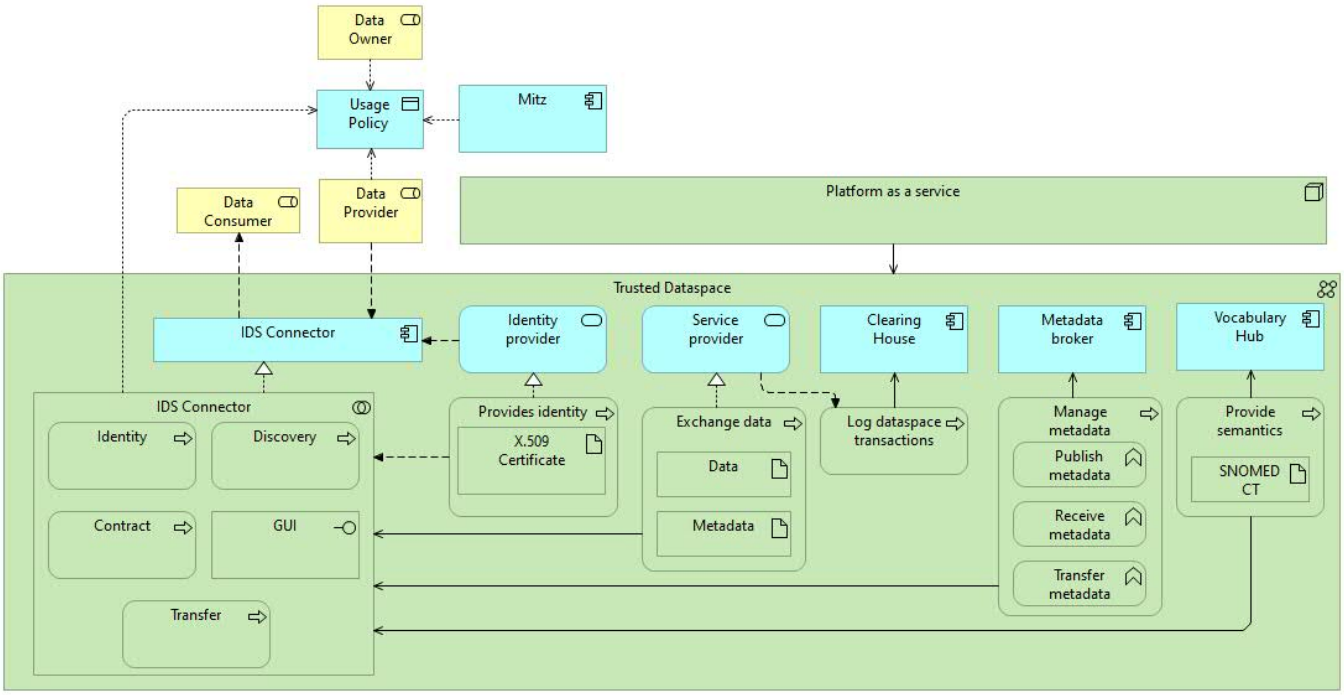
- Would you say this system is capable of correctly registering patient data sharing consent?
- Would you say this model is secure?
- Would you say this model is scalable?
- Would you say anything is missing or incorrect in this model?
- Is there anything you would like to comment on not mentioned earlier?

- Show IDS Connector model



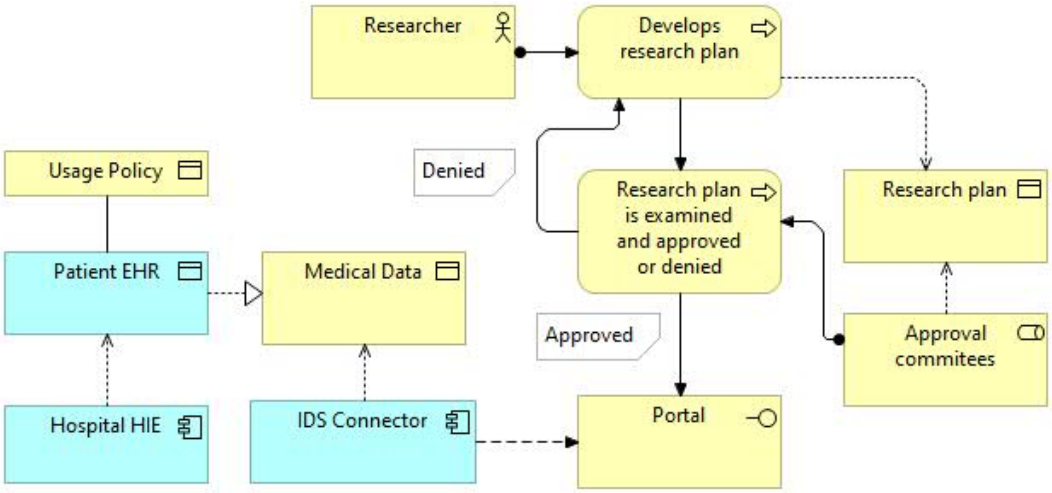
- Would you say this system is capable of providing the functions required of a data connector in a dataspace?
- Would you say this system provides the data transformation requirements of a medical dataspace?
- Would you say this model is secure?
- Would you say this model is scalable?
- Would you say anything is missing or incorrect in this model?
- Is there anything you would like to comment on not mentioned earlier?

- Show Technology model



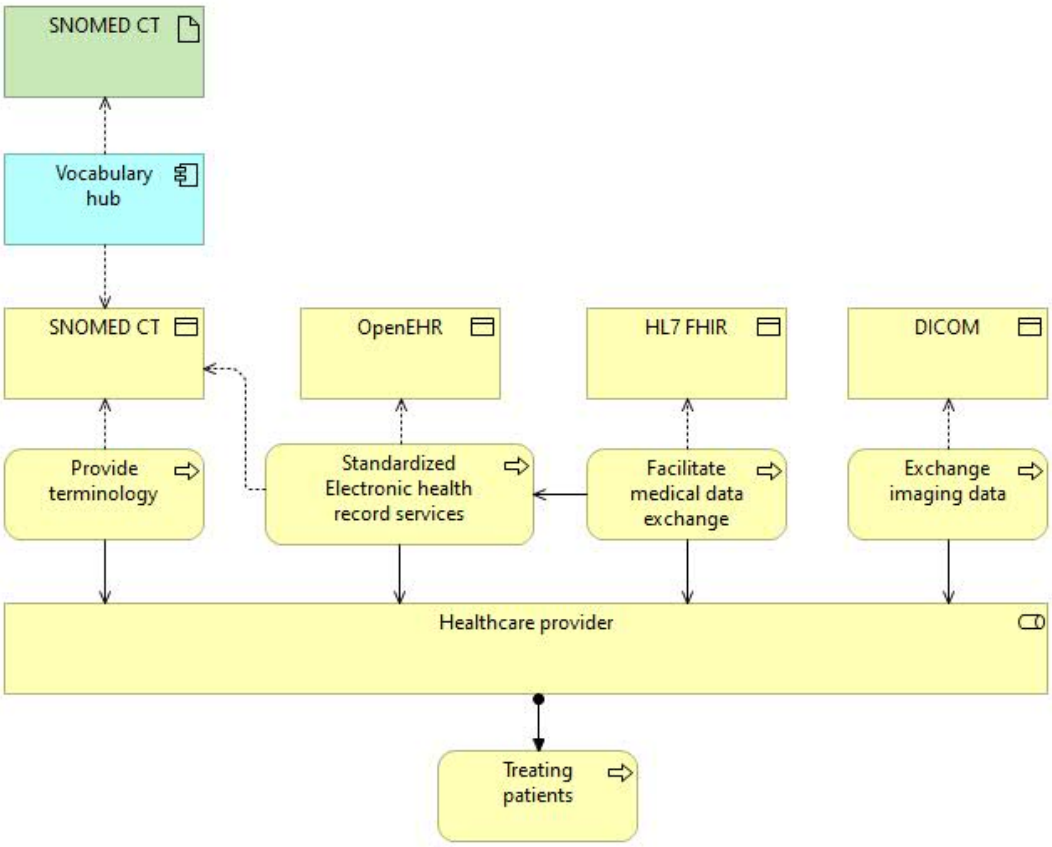
- Would you say this system correctly supports the application components present in the previous steps?
- Would you say this model is secure?
- Would you say this model is scalable?
- Would you say anything missing or is incorrect in this model?
- Is there anything you would like to comment on not mentioned earlier?

- Show researcher model



- Would you say this system correctly describes the most basic process for getting a research plan approved and getting access to medical data?
- Would you say the application elements correctly provide the required functionality for retrieving the data for researchers?
- Would you say this model is secure?
- Would you say this model is scalable?
- Would you say anything is missing or incorrect in this model?
- Is there anything you would like to comment on not mentioned earlier?

- Show Standards model



- These standards represent the most common standards in the dataspace. Would you say these are represented correctly in this model?
- Are there any standards you are missing?
- Is there anything you would like to comment on not mentioned earlier?

- Show dataspace overview model

- Now that you have seen the more detailed elements of this dataspace. Do you see any potential bottlenecks?
- Would you say the architecture as a whole is scalable?
- Would you say the architecture aids interoperability between healthcare providers?
- Would you have any other comments about the overview model not previously mentioned.

The following questions were asked in the two interviews with the MST architects to represent stakeholders in the Twente region. These were asked to validate the problems identified during problem the identification phase.

Lack of regional medical data exchange:

- How would you describe the current state of regional medical data exchange in your experience?
- How frequently do you find that regional data exchange issues hinder collaboration between healthcare facilities?
- Would you describe the current situation regarding regional data exchange as a problem?

Need for interoperability and data sovereignty

- How important is interoperability between different healthcare systems and platforms?
- Do you feel patients currently experience data sovereignty?
- Do you feel their is progress to be made in terms of interoperability between systems and organisations? What about in terms of data sovereignty?
- Would you say that interoperability is a current issue in the medical field?

Patient Privacy and Regulatory Compliance

- How do you currently ensure that patient data privacy is maintained while meeting regulatory compliance standards in your architectural projects?
- How do you think current regulations (such as GDPR) impact the design and implementation of data systems within the medical sector?
- Do you feel that regulations are limiting interoperability in the medical field?

Appendix C

Appendix C - Expert Validation Interview conclusions per interview, per model

1. **Cloud Solution Architect:**
   - **Initial look at dataspace model**
     - This model does not feature any checks on data quality.
     - There is no mention of what happens when you add more connectors (more than 2 in this case).
   - **Consent Model:**
     - The expert would expect a mechanism to see who has access to patient data. While Mitz does seem to offer this functionality it is not specifically mentioned in the model.
     - Mentioned that it does not mention any updates to the metadata broker when Usage Policy is updated later. Given this is personal data they felt it is important to have it mentioned.
     - The expert suggested generalizing the model, not specifically mentioning Mitz as this could get outdated with time if Mitz is no longer the accepted standard.
   - **Connector Model:**
     - Wondered why the data transformation was done in the connector and not prior, more centrally. This could avoid having to do the same data transformation repeatedly.
     - Security not mentioned in the model. Would expect increased security on all incoming and outgoing connections and TLS certificates in place to support this.
     - The scalability concerns related to the aggregating of data transformation mentioned above.
     - Questioned the use of docker (containers) for the IDS connector.
     - Questioned whether I had any expectations regarding downtime of the system given the critical nature of a medical data system.
   - **Technology layer model:**

  - The expert felt that the system does support the application components correctly.
  - Because a PaaS solution keeps data away from the source, this forms an inherent risk.
  - In general, felt that the model did not contain adequate information to determine security of the model.
  - Similar to previous points, doubted whether the model is scalable when many requests are made at the same time.
- **Researcher viewpoint:**
  - Felt that the displayed process for requesting access to the secondary use of data was correct.
  - Felt that the application components correctly provide the functionality required for the secondary use of data.
  - Felt it was difficult to make a judgement on the scalability or security due to lack of information.
  - Felt the dataspace should have a functionality which tracks the research projects utilize the dataspace for future reference.
- **Standards Model:**
  - Skipped due to time constraints.
- **Repeat of dataspace model:**
  - Expert was surprised that the clearing house is not connected to the metadata broker.
  - Found it hard to determine the scalability without know more about the context, like number of users.
  - The expert expressed worries about what would happen if many connectors request the same data element at the same time.

2. **Technical Lead:**

- **Initial look at dataspace model**
  - Expert was very surprised at the use of the term data owner, personally only uses the terms present in the GDPR, Data Subject, Data Controller and Data Processor.
  - Missing De-identifcation (anonymization/pseudonymization) in the main model.
  - For the primary use of medical data would not expect to see a data catalogue like the metadata broker, instead would expect some form of traditional data localization to determine where patient data is being stored.
  - Does not see the data request feature in the dataspace.
  - Mentioned that in the current Dutch medical landscape there is a movement towards FHIR end points.
- **Consent Model:**
  - Felt the model can correctly register patient data exchange consent.
  - Answered that the model can add to data security, assuming that every time that the data is shared the usage policy is consulted to ensure the exchange is authorized. The logging of data exchanges can also aid in the transparency and thus, trust in the system.

  - Felt the model is crucial to the scalability of the dataspace because all healthcare providers logging data sharing consent local is not scalable.
  - Felt the model could be more generic by not using Mitz specifically but a Zeggenschapsregister (Authorization registry in English)
  - Felt that the secondary use of data was missing from the model.
  - Not directly related to the model: Expert prefers opt out over the current system of opt in to sharing medical data.
- **Connector Model:**
  - The expert expressed he would expect to see a FAIR data point in the connector, especially for describing datasets for the secondary use of medical data. This is a data point which enables FAIR data compliant data sharing.
  - Recommended changing the Data Anonymization process to Data De-identification as to also allow for data psuedonymization.
  - Mentioned a current point of discussion with regards to the viewing of patient EHRs. The expert talks about how it is important to have different versions of EHRs through time, to ensure that it can be determined what an EHR looked like at the time of it being requested by a healthcare professional. This is currently not present in the architecture.
  - Questioned the difference between the catalogs functionality in the top left of the model, versus the metadata broker application at the bottom.
  - Expert was not sure whether the connector would cause problems for scalability as it would create a single point of failure for the system if something happens to it.
- **Technology layer model:**
  - Reiterated his doubts about using data owner in the models.
  - Thinks the system should correctly support the application components present in previous models.
  - Due to the overlap between the technology model and the previous models, the expert had no new insights regarding scalability and security.
  - The expert is missing a data centric model.
- **Researcher viewpoint:**
  - Felt the right side correctly represents the research journey from the top of their head. Additional information could be found at the health-RI website.
  - Missed a data approval committee also on the left side of the model (which is the data/healthcare provider side).
  - Felt there were several element which could make the portal element more detailed.
  - After looking at the IDS connector earlier he does not feel that this model could work as is, without further development on the FHIR endpoints.
  - Felt the process is scalable, if the extra detail described above is added.
  - Felt the model is generally secure, if a secure process environment is included and the data stays where it is allowed to be stored.
- **Standards Model:**
  - Skipped due to time constraints.
- **Repeat of dataspace model:**

- Felt the brokers could cause a bottleneck in the system.
- Felt the system as a whole is scalable, outside the points raised during discussion of earlier models.
- Generally the division between primary use of data and secondary use of data could be made more explicit.

3. **Manager Data & Information:**

   - **Initial look at dataspace model**
     - Interviewee was already familiar with the dataspaces concept.
     - Felt the model could use clarification on the process flows.
   - **Consent Model:**
     - Felt the system would be capable of logging consent correctly
     - Mentioned that the model does not contain adequate information to make a comment on security. This is a conclusion that was common throughout the models.
     - Felt they didn't have enough information to determine whether the model is scalable.
   - **Connector Model:**
     - Felt the connector would be able to fulfill the functionalities required of the connector.
     - Felt it also is capable of performing the required data transformations.
     - Felt they had too little information to make an appropriate comment on the safety of the model.
     - Felt the model is certainly scalable
   - **Technology layer model:**
     - While they felt the required components were there, they felt a lack of security was an issue that had to be resolved before making a definite decision on this topic.
     - Felt the model is scalable.
   - **Researcher viewpoint:**
     - Felt the model appropriately described the process for getting research approved.
     - Felt the application components appropriately covered the needs of the system.
     - Felt there was not enough information to form an opinion on security.
     - Hard to determine scalability.
   - **Standards Model:**
     - Due to a sudden event on their end we had to shorten the meeting and thus this model was skipped.
   - **Repeat of dataspace model:**
     - When asked about bottlenecks they mentioned that data governance might be an issue limiting the development of a medical dataspace.
     - Felt that due to the limited context in the models it is hard to determine. Suggested starting with a proof of concept with two exchanging parties.

- Felt the model does enable interoperability of medical data between healthcare providers.

4. **Lead Architect:**

   - **Initial look at dataspace model**
     - At first look, did not have any comments regarding the dataspace model.
   - **Consent Model:**
     - The personal opinion of the expert does not align with the way Mitz operates. The expert feels that consent should be based on a healthcare professionals role, not the organisation they work for.
     - In their reply to the security question they referred back to the point above. They do not feel the model is secure when consent is given based on what organisations can use your data instead of what healthcare related role someone fulfills.
     - Felt the model was "endlessly scalable".
     - Felt a final check of the consent was missing.
   - **Connector Model:**
     - Felt the connector contains all functions required of a connector in a medical dataspace.
     - Misses a check for data quality on incoming data.
     - Felt the model does not contain enough information to form a opinion on security of the model
     - Felt the model was scalable, partly due to containerization with Docker. However, this can be achieved in other ways and thus wondered why docker was being used here.
     - Questioned why hardware was included in the model.
   - **Technology layer model:**
     - Reiterated here how he misses a check for data quality.
     - Could not form an opinion on security due to lack of information in the model.
     - Felt the model was scalable.
   - **Researcher viewpoint:**
     - Felt the process flow to get access to medical data depends on the healthcare organisation.
     - Misses an element between portal and the connector, which would realize the portal.
     - Felt it hard to form an opinion on the security of the model due to it being very function focused.
     - Did mention that for personal data security it was detrimental that the usage policy is centrally stored in a secure manner.
   - **Standards Model:**
     - Felt the standards were represented correctly in the model.
     - Expert felt no standards were missing.
   - **Repeat of dataspace model:**

- Worried about what happens when a lot of connectors get added and the amount of data grows with that.
- Other than that does deem it scalable.

5. **Cloud and Infra Architect:**

- **Initial look at dataspace model**
  - Did not have any initial thoughts on the dataspace model.
- **Consent Model:**
  - The model should be adequate to register consent, given it represents Mitz correctly because Mitz is adequate.
  - Expert could not provide an opinion on security as not enough information is present in the model.
  - Assuming Mitz is scalable the expert felt this model is scalable too.
  - Felt the Mitz portal should be represented using different elements.
- **Connector Model:**
  - Felt a component was missing to represent temporary data storage within the model.
  - Felt that the data transformation processes could be simply another function to add to the Connector Functionalities component.
  - Expert could not provide an opinion on security as not enough information is present in the model.
  - In terms of security, would want specifically privacy be design/security by design to be represented.
  - Felt the model is flexible
- **Technology layer model:**
  - Felt the application components are supported correctly as that is simply a design decision.
  - Expert could not provide an opinion on security as not enough information is present in the model.
  - Expert questioned the use of a communication network to represent the trusted dataspace.
  - Expert liked the use of the X.509 certificate for identity verification.
- **Researcher viewpoint:**
  - Expert would like to see role based access in the model.
  - Related to the point above, the expert would prefer to see an extra check whether a researcher has the required authorization to access data before actually receiving the data.
  - Expert could not provide an opinion on security as not enough information is present in the model. However, did reiterate the importance of role based access when asked about security.
- **Standards Model:**
  - Felt the standards were represented correctly in the model.
  - When asked about missing standards they suggested adding other HL7 standards.

- **Repeat of dataspace model:**
  - Felt no bottlenecks are present in the architecture.
  - Felt the system was scalable because all underlying components were generally scalable.

Next, as the two architects whom work at MST are also considered stakeholders, they were asked additional questions to verify the identified problems. The questions can be found in Appendix B.

1. **Lead Architect:**

   - **Lack of regional medical data exchange:**
     - When asked about the current state of regional medical data exchange he said, quote: 'We fax'.
     - Mentioned there is currently no centralized data exchange platform.
     - When asked about daily problems regarding interoperability they mentioned that anyone who requires urgent care could have a problem with their data not being available to healthcare professionals in time.
     - When asked whether it is a problem, he said that one could raise it to be an issue. That said, he personally does not necessarily see it as that because it is only an issue because recent developments allow for it to be.
   - **Need for interoperability and data sovereignty:**
     - When asked about the importance of interoperability they mentioned that standardization is the most crucial challenge.
     - Did not feel that patients currently experience data sovereignty.
     - Felt improvements were possible in both interoperability and data sovereignty.
     - Again, felt this is currently not an issue because modern developments only allowed it to become an issue where it was previously not a problem.
   - **Patient Privacy and Regulatory Compliance:**
     - Ensured GDPR compliance through ensuring data is stored safely in their work. When that data is being used it is no longer in the control of an architect.
     - Felt there are two ways in which the GDPR influences design and implementation of systems. The first is how data is stored, and the second is how it is being used. The first they can control, and they do. The second, it out of their domain.
     - Does not feel regulation limits interoperability in the healthcare sector. Even felt it promotes it as encourages collaboration, for example through Wegiz.

2. **Cloud and Infra Architect:**

   - **Lack of regional medical data exchange:**
     - Expert felt that the current state of interoperability in the region is poor to bad.
     - In his daily job activities they do not run into issues regarding interoperability, but improvement is possible.

  - When asked whether interoperability is a problem, they responded that it is not being worked at for nothing.

- **Need for interoperability and data sovereignty:**
  - The stakeholder considered interoperability between healthcare systems and platforms very important.
  - Felt that patients experience data sovereignty when everything goes according to its design.
  - Felt improvements can be made in terms of both inter-system interoperability and patient data sovereignty.
  - Considered interoperability between systems a problem in the healthcare sector, mostly due to problems with standardization.
- **Patient Privacy and Regulatory Compliance:**
  - Patient privacy and regulatory compliance in their work is ensured through the use of two security officers and the following of NEN7510.
  - Their work is influenced in the sense that when designing systems, they always follow privacy/security by design principles.
  - Does not feel that regulation limits interoperability. However, it is something to keep into consideration and adds a challenge.

Appendix D

Appendix D - Requirements and their fulfillment

1. **There needs to be an ontology and common vocabulary in place in order to ensure semantic interoperability:** *Fulfilled*
   The dataspace uses SNOMED CT to ensure semantic interoperability.

2. **Negotiation of data exchange needs to be automated:** *Fulfilled*
   Through the usage policies and the use of the broker and service provider the negotiation of the data exchange can be seen as automated.

3. **Usage contract creation needs to be facilitated through a graphical user interface:** *Fulfilled*
   Mitz features an online graphical user interface.

4. **There needs to be the ability to transform data to ensure interoperability, and in doing so, maintain interoperability between standards:** *Fulfilled*
   The architecture of the connector in this dataspace features a data transformation element to ensure interoperability between standards.

5. **Usage policies need to be enforced:** *Partially fulfilled*
   The enforcement of usage policies is not explicitly mentioned in the architecture design. However, organisations that ignore usage policies would face (legal) consequences.

6. **Data needs to be described using metadata, allowing for discovery through metadata in the dataspace:** *Partially fulfilled*
   HIE (Health Information Exchange) systems which is what the healthcare providers use to log information typically also describe their data using metadata. However, given that this is up to their developers it cannot be guaranteed. That being said, the system MST uses (Hix) does feature the process of describing data using metadata and the dataspace features metadata discovery.

7. **In case of malfunction of the dataspace, the relevant stakeholder must be notified automatically:** *Fulfilled*
   The dataspace architecture features a system for automatically notifying IT staff in case of problems.

8. **Upon completion of the dataspace, a legal and organizational compliance review needs to be completed to ensure compliance with all EU and na-**

tional legislation: *Not fulfilled*
The system itself does not feature a legal compliance review. However, this can be executed in the future if required.

9. **An overview of participants of the dataspace must be available:** *Not fulfilled*
The system currently does not have a comprehensive list of all participants in the dataspace. This could be added externally in the future.

10. **A user needs to be able to access all relevant data and services, using a single set of credentials:** *Fulfilled*
The only credentials a patient needs to operate the system is through Mitz (which uses DigiD, the Dutch government digital identity verification system). Healthcare providers should be able to still use the login credentials they currently use for their HIE system.

11. **A ticketing system is to be provided for users to report technical issues or get support:** *Fulfilled*
As shown in figure 4.16, this is facilitated in the dataspace.

12. **Participants in the dataspace retain data sovereignty:** *Fulfilled*
This is inherently true about any IDS-RAM adhering dataspace as it is a core value. It is realized here through the Mitz system interaction.

13. **The dataspace needs to have defined policies which specify what attributes an applicant must have to become a trusted participant in the dataspace:** *Not fulfilled*
This is up to the participating organisations and ZorgNetOost to determine. Additionally, this requires legal approval as to not violate any existing legislation.

14. **The architecture must ensure that the healthcare provider needs to be able to have access to their patients' information when required:** *Fulfilled*
This is a core value of the dataspace and the architecture is designed with real-time data access in mind.

15. **There need to be security (authentication and authorization) measures in place for participants:** *Fulfilled*
The IDS-RAM and this architecture by extension are designed with security in mind. This is exemplified through the use of proven HIE systems and IDS connector design based on models which were produced with a heavy focus on security from TNO and Sovity [101].

16. **Processes need to be described on how to onboard participants, offer data, perform data contract negotiation, exchange data and publish using data apps:** *Not fulfilled*
This is currently not created and would have to be done prior to practical implementation.

17. **Healthcare providers need to be able to use their current software:** *Fulfilled*
In the current design of the dataspace healthcare providers would keep utilizing their current HIE systems which in turn exchange that data with an IDS connector.

18. **Patients need to be able to determine themselves what parties can share and receive data:** *Fulfilled*
This is facilitated through the Mitz $\leftrightarrow$ usage policy interaction.

19. **Usage policies need to be able to be changed in real time:** *Fulfilled*
Mitz features real-time changes, given that the dataspace uses their information to create the usage policies, the usage policies can be changed real-time too.

20. **Updates to patient data cannot take longer than 5 minutes to be available to healthcare workers:** *Partially fulfilled*
The architecture aims to facilitate real-time updates of patient data. However, the actual update speed depends on the technical implementation and the performance of the involved systems.

21. **HL7 FHIR needs to be used:** *Fulfilled*
Given various reasons provided throughout this thesis, HL7 FHIR is the standard used for medical data exchange in this dataspace.

22. **Data collected upon first entry into the hospital needs to be available throughout the process:** *Partially fulfilled*
The design promotes the availability of data throughout the care process. However, the actual availability depends on the implementation of data exchange between healthcare providers and the systems used.

23. **Lab data must be available to primary care providers (where authorized):** *Fulfilled*
As long as the lab data is entered into the HIE system, the primary care providers should have access to the data too.

24. **The system needs to be accessible by both phone and PC:** *Partially fulfilled*
As long as the HIE system which a healthcare provider uses supports phone and PC utilization, so should the dataspace as the IDS connector is connected to existing systems

25. **Nanda NOC NIC or SNOMED CT is to be used:** *Fulfilled*
SNOMED CT is used as mentioned above.

26. **Retrospective data for research needs to be available (where authorized):** *Partially fulfilled*
Similar to how requirement 24 was dependent on the HIE system used, so is it here. If the system used to access medical data stores the information long enough for retrospective research, so will the dataspace.

27. **There needs to be a check in place to ensure researchers have completed all required paperwork before being allowed to access data in the system:** *Partially fulfilled*
In the researcher viewpoint in figure 4.12 there is a check on whether the paperwork is completed. That being said, this check takes place outside of the dataspace in the preparatory phase before gaining access to it. Thus, this requirement is deemed partially fulfilled.

28. **There needs to be an anonymization process available:** *Fulfilled*
The design of the data connector contains a service which allows the data to be anonymized.

29. **Patients need to be able to opt out of their data being used for research purposes:** *Fulfilled*
This is achieved in the Mitz system.

Above, 29 requirements are presented. Of these 18 are fulfilled, 7 are partially fulfilled and 4 are not fulfilled.