Pandemic Pandemonium in de GGD: towards FAIR Data in Health

MASTER'S THESIS VERSION FINAL
ICT in Business and the Public Sector &
Computer Science: Data Science
GGD Corona systems - Case Study

Leendert van der Plas (s2049600)

Date:
28/08/2023

Word Count:
119635 words

First Supervisor:
Prof. Dr. Mirjam van Reisen of the Leiden University Medical Centre

Second Supervisors:
Drs. Peter van Veen of the Leiden Institute of Advanced Computer Science (ICT in BS. & PS.)
Dr. Katy Wolstencroft of the Leiden Institute of Advanced Computer Science (CS: Data Science)

Key Words

AcICT, Adviescollege ICT Toetsing, AP, Autoriteit Persoonsgegevens, AVG, Common Standard, Data Science, Data Visitation, Data Protection Impact Assessment, FAIR, Framework, GDPR, GGD, Healthcare, ICT, IT, Netherlands, Open Source, Privacy, Scientific Research, Security

Acknowledgment

I would like to express my deepest gratitude to my first supervisor, Prof. Dr. Mirjam van Reisen, for her invaluable guidance and support throughout both of my master's programs and for connecting me to a variety of experts. I would also like to extend my heartfelt appreciation to Drs. Peter van Veen, who provided crucial assistance on the thesis section for my first master's program, and Dr. Katy Wolstencroft, who offered her expertise for the thesis section of my second master's program.

For the successful completion of this thesis, I conducted numerous in-depth interviews and informal conversations with individuals actively involved in the Dutch government. I am grateful to all of the participants for their time and valuable input. In particular, I would like to acknowledge the following individuals: Ron Roozendaal, the former CIO of the Ministry of Health, Welfare, and Sport (VWS) during the onset of the Covid-19 pandemic and the Deputy General of Digitalization in 2022; Mustafa Kedilioglu, a Project and Program Manager at the Ministry of VWS; Ronald Damhof, the current Chief Data Officer of the Ministry of Justice; and two anonymous management-level individuals at the GGD who generously shared internal information about the organization.

Foreword

This thesis has been created for the masters of ICT in Business and the Public Sector and Computer Science: Data science, with Sections 5 and 6 corresponding to the first masters and Sections 7 and 8 corresponding to the second masters. These masters, in addition to my bachelor's in Computer Science and Economics and my third master's in Crisis and Security Management: Cybersecurity Governance offer me an extensive background in applying ICT in organizations and addressing security threats through technical and organizational measures, but do not offer a legal background.

The legal aspect of this thesis has been created based on interviews with a Data Protection Officer employed at one of the Dutch ministries, a legal advisor with an extensive background in the GDPR, other governmental officials, and an investigation into the GDPR and the surrounding laws related to processing medical information. This thesis clearly discusses the findings from my investigation and its probable violations of the GDPR based on this, but only the Autoriteit Persoonsgegevens or a judge can determine this conclusively. Although the Autoriteit Persoonsgegevens and the currently ongoing mass claim from Stichting ICAM offer strong evidence of the GDPR being violated.

This thesis offers a very comprehensive investigation into the problems occurring at the GGD during the covid-19 pandemic, with information being presented as it has been presented to me from governmental reports, various interviews I have conducted, and any other information I was able to gather. While some of the information in this thesis has come from internal interviews with governmental personnel involved at the GGD and beyond, the majority of the information has come from publicly available sources. Additionally, at no point during the creation of this thesis was I influenced by any such organizations in any way. Neither is it the intention of this thesis to appear as some kind of political reckoning, with any negative claims being supported by evidence and only functioning to identify and investigate problems that led to a possible lack of GDPR compliance.

Given the fact that GGD GHOR does not fall under the WOO, although each individual GGD does, the WOO was only of limited use in gathering additional information. Therefore, while this has been a very comprehensive investigation, only the Dutch governmental can fully and completely investigate the findings from this report and any additional problems that were inaccessible for this thesis. Which would most likely be done through a future parliamentary inquiry.

In the creation of this FAIR-based framework, I have focused on increasing the level of transparency at every stage of project development, as I believe this to be critical to the success of any project. Previous governmental initiatives such as the ‘beleidslijn ‘open, tenzij’’ and the algorithm register follow this same line of thinking although this framework takes these a step further. This increase in transparency has three main goals, being to allow the detection and use of patterns, to allow individuals to contribute and address the lack of expertise among IT employees, and the final one being to (partly) reverse the effect of political influence. Transparency makes it unwise to make any decision that would be viewed negatively or ignore any recommendation or standards.

For the proposed FAIR-based framework, it is important to consider that the safety, security, and success of any project is not reliant on its use of technology alone. Current technology can construct systems that are safe and secure, yet if organizations are either unwilling or unable to monitor their development or make incorrect decisions at any point, it is unlikely to result in anything but an insecure system. The proposed FAIR-based framework, as presented in this thesis, should be able to result in GDPR-compliant projects, but only when correct organizational decisions are made to remain in compliance. If this concept is not followed, any approach that intends to fulfill any of the government’s initiatives, including the proposed FAIR-based Framework, will not be successful.

Abstract

This study investigates the problems of compliance with the General Data Protection Regulation in health data systems run in the Netherlands since the GDPR went into effect on the 25th of May 2018. The focus of this investigation will be on three systems, (HPZone, HPZone Lite, and CoronIT) used by the Dutch municipal health services (GGD) during the Covid-19 Pandemic (January 2020-2023). It is probable that these systems violated numerous articles of the GDPR as well as the basic principles of 'data minimalization' and 'privacy and security by design and by default'. The design of these systems was unsafe, with significant numbers of untrained and unmonitored employees being given broad access without appropriate limitations and safeguards. Procedures through which these problems could have been identified were also not carried out properly. The main contributors to these problems are the application of previous knowledge in an inappropriate context, an unmanageable growth rate, the presence of external organizations combined with a lack of control, and a high level of political involvement in what should have been solely a healthcare crisis. The federated structure of the GGD as well as its unique position in the Dutch government also complicated matters.

Based on this investigation into the GGD, relevant background information relating to his project and the wider state of governmental ICT, conferences with attending high-level governmental employees, and additional interviews with a project manager and the Deputy General of Digitalization, the specifications and requirements of a GDPR compliant healthcare ICT support system were determined. With a focus on systems used by a larger number of employees with a lower level of expertise, as this situation is the most comparable with the GGD case study and poses the most significant risk to the privacy and security of the data subject. These requirements and specifications were then compared with the FAIR guiding principles and relevant approaches to data management to determine which of these aspects could be addressed.

Based on this information, a proposed FAIR-based framework was created based primarily on the FAIR guiding principles, data federation, and data visitation. The framework itself can be divided into three main areas of improvement. The first area consists of the content and design of a Dutch governmental ontology spanning from the organizational layer down to the variables used in various projects and systems. The second area consists of limiting the amount of information being exposed to employees based on what's required for their role, adhering to the previously mentioned GDPR principles. This information is included at the project/system variable level, either as requirements or recommendations, to shape system design. The third consists of a common base through which projects can be developed and deployed, which is critical given the lack of experience with FAIR in organizations. Based on this approach, there is always a deployable and working base, which can then be adjusted with project-specific implementations. Continuing with this approach, every project built based on the FAIR framework could be used as an alternative starting point instead of the default base, allowing for either near-immediate deployment of federated systems or reducing the amount of development required to create different but comparable systems.

To answer the Research question 'To what extent could FAIR guidelines improve governmental healthcare systems' compliance with the GDPR in a crisis situation?', each of the main areas of improvement can aid in improving GDPR compliance. However, it is impossible to guarantee that all systems using this approach will fully adhere to the GDPR, given that this is dependent on project-specific technical and organizational measures. Additionally, to be fully effective in a crisis situation, this approach must be widely used. Only through constant use can the government gain the required level of expertise and information, and alternative projects and starting points.

Table of Content

List of Figures

List of Tables

1. Introduction

1.1 Literature Review

It is becoming increasingly important that health and social care services work as a single unit, compared to each organization having different priorities and standards. This aids in giving patients a seamless care experience, reduces costs as system development is shared instead of parallelized, and ensures that medical professionals have access to all information [1, p. 31]. However, as the amount of information that is accessible, via a single system, increases, the potential for abuse also increases [2]. A single personal data breach may now compromise the entire repository of personal data, instead of a limited portion of siloed information.

The security and privacy of personal data play a vital role when healthcare is supported by ICT, based on the findings of a study by Wiktora and Martina about the perceived importance of various user requirements [3]. Security refers to securely storing and transferring data, to guarantee its integrity, validity, and authenticity. Privacy refers to the fact that data may only be accessed by people who have the authorization and need to view and use it [4, p. 1]. As people require and insist on the highest security and privacy standards, any healthcare system needs to take these aspects into account. While this would suggest that users are able to make informed decisions about sharing their personal data, the privacy paradox suggests that this does not happen. Based on this paradox, while people highly value their fundamental right to privacy, they do not act accordingly, especially when it concerns new technologies [5]. With the vast majority of people never reading privacy policies at all [5].

While the findings from a study by Wiktora and Martina measure perceived importance instead of actual importance, this perception can lead to individuals and patients avoiding doctors and the rest of the healthcare system because they are afraid of their data not being secure [1, p. 33]. As data breaches occur relatively often (statistics of personal data breaches are discussed in Section 4.3.2), this fear is not entirely unjustified. Importantly, an individual’s worries about security issues could lead to public health as a whole being compromised, for example during a pandemic. In such situations, people must get tested and/or vaccinated, which requires offering up their data.

Data protection in the European Union (hence ‘EU’), has initially been primarily up to the member states in question, although based on a common framework. This data protection can be divided into three main phases, starting in the 1980s with national personal data protection laws being adopted based on the Organization for Economic Co-operation and Development (OECD) guidelines [6] of 1980 and a Council of Europe convention [7] in 1981 [8, p. 3]. For example, in the UK this resulted in the Data Protection Act of 1984. This act gave new legal rights to individuals whose personal information was stored on a computer, aiming to regulate the use of automatically processed information relating to individuals and the provision of services in respect of such information [9]. Although other countries differ in the exact implementation, the general principle remains.

The second phase was based on the European Data Protection Directive of 1995 [10]. Given that it was a directive, it only specified the results that must be achieved instead of dictating the exact manner in which the regulation would have to be achieved. Based on the Data Protection Act (hence ‘DPA’) of 1998 in the UK, this act introduced new provisions for the regulation of the processing of information relating to individuals, including the obtaining, holding, use, or disclosure of such information [11]. The act also expanded on the concepts of digital media and computers as their use had significantly increased since the creation of the previous legislation.

The third phase is based on the General Data Protection Regulation [5] (hence 'GDRP'), which aims to give people more control over their private information and to mitigate abuse of personal data in the medical sector or any other sector from happening. It applies to any interaction with EU citizens, even if the organization in question operates outside of the EU (Article 3 of the GDPR). Due to the difficulty of separating parties, it's often standardized so that organizations work in compliance with the GDPR for all parties. The GDPR and its implications will be discussed in more detail in Section 2.1.

Given that the GDPR went into effect on the 25th of May 2018, this is a relatively recent topic and would limit the amount of material that could be used for this thesis. However, as the GDPR builds upon the previous two stages, the security and privacy protection measures from an earlier date are still valuable for this thesis, as long as the GDPR takes precedence.

Some studies have been done on the safety of systems, in the context of the GDPR, such as an evaluation of GDPR compliance in Mobile Health applications [11]. A study by Brodin identifies problem areas with the GDPR and incorporates this into a framework [12] based on his earlier work from 2015 [13]. However, this framework is limited in scope. It offers only a global view of the analysis and design phase and lacks to mention nor focus on healthcare and/or governmental organizations. Therefore, there seems to be no framework for implementing GDPR compliance in any governmental healthcare system to use right now. To fully comply with the GDPR, such mechanisms, certifications, and frameworks must be created.

A study by Shu and Jahankhani evaluated the impact and implications of the GDPR for a governmental healthcare system [14]. In this study, they contrast the previously acting DPA from 1998 with the new GDPR act of 2018, in the context of the English National Health Service. Major differences between the 1998 DPA and the GDPR are that "data protection must not only be by default but must be by design (Privacy by design)" (GDPR Article 25), the need for every organization to respond to requests by the data subject within 1 month at the risk of fines (GDPR Article 15) which comes at great organizational difficulty and cost and the fact that any breach of personal information now needs to be reported within three days to the supervisory authority that governs the GDPR in its respective Member State (GDPR Article 33). While Britain has since left the EU, this study and analysis of the implications of the GDPR were from before that, meaning that the changes described in this study should apply to any other governmental healthcare system under the GDPR.

Furthermore, Shu and Jahankhani suggest that this data protection act is also an opportunity for different organizations to share information about potential risks regarding the processing of personal data, which could be incorporated into models to aid other organizations. It is also "unrealistic given the nature of some of the new requirements that the GDPR will bring about, that a lot of smaller scale healthcare providing organizations will be able to meet these requirements acting alone, or without significant advice, support and guidance" [14, p. 37], which will likely apply to any other country too. Further highlighting the need for practical implementations of GDPR-compliant systems instead of lists of requirements.

The current focus of academic research in improving the privacy and security of healthcare data seems to be related to the blockchain [15] [16]. Tapscott describes blockchain technology as "especially attractive to the healthcare industry due to the interoperability and security features that can create a mechanism by which personal data can be anonymized and transferred securely to different medical units and research centers, reducing time and cost" [17]. This suggests that this technology could be used to improve GDPR compliance.

However, a study by Manatunga, comparing the legal requirements of the GDPR with the blockchain, determined that the very principle of the blockchain is incompatible with the GDPR [18]. Following Article 17 of the GDPR, a data subject has the right to erasure, but given that the blockchain is immutable, it is not possible to erase anything [18, pp. 36-37]. Data is also distributed over the entire network, making it difficult to impossible to provide accountability and responsibility [18, pp. 39-40]. And as this is a key characteristic of blockchain technology, a newer version of the GDPR would have to be created to be able to comply. This is possible as the acting data protection act has been changed numerous times to incorporate new technology, but makes it impossible to use right now [18, p. 49].

1.2 Problem Statement

For personal data, and especially for personal health-related data, it is of significant importance that this data remains both private and secure. As healthcare systems have become bigger in both the number of data subjects contained in these systems and in the amount of information about each data subject, it has become even more important to ensure that data remains well protected. In the healthcare sector, even the perception of privacy and security is important as any situation where individuals are unwilling to interact with the healthcare system can result in real damage to public health.

The GDPR, the latest iteration of the European data protection laws, aims to achieve that data remains private and secure through its core tenets of “privacy by design and by default” and “data minimalization”, meaning that the least amount of data is being collected for a given purpose and that privacy and security are taken into consideration at every step in the process of system design and implementation. However, implementing the GDPR correctly remains an issue in many organizations and governmental ICT projects, and ICT projects in general, are often unsuccessful [19] [20]. Therefore, just the existence of the GDPR has failed to ensure that personal data remains private and secure, even with the threat of 9-figure fines for the worst offenders [21].

This thesis looks at a more specific and more difficult scenario, concerning GDPR compliance in the use of governmental healthcare systems during a crisis situation. While no crisis is alike, common characteristics are a significant amount of time pressure, a large need for resources, and a lack of preparedness for the situation. Developing and implementing projects in such a scenario is difficult, and adhering to the GDPR to ensure that personal data remains both secure and private is even more difficult. Therefore, a project in this scenario can only succeed when working with a clear and implementable framework, to ensure that GDPR compliance is maintained.

1.3 Research Gap

Current research indicates that it remains difficult for many organizations to ensure GDPR compliance in a normal situation, let alone a crisis situation where there is a lack of time and experience to develop projects. Yet it is in these crisis situations that it is critical that projects are developed correctly and ensure that data remains private and secure. Given that the GDPR act went into effect as recently as the 25th of May 2018, a limited amount of academic research has been done on GDPR compliance in governmental healthcare systems in general. Research identifies areas that need more attention with the GDPR than with the previous DPA act, but does not offer any solution on how to be compliant.

This thesis will fill in this research gap by aiming to be this solution, through developing an architectural framework that can be used in a crisis situation in the healthcare sector, and by extension in a non-crisis situation too. Currently proposed solutions like the blockchain are both too abstract to serve as a framework and not suitable for implementation until some point in the future. Instead, this architecture will apply the FAIR guiding principles, an already existing scientific standard towards data management, to areas lacking GDPR compliance as identified in this investigation.

1.4 Research Objectives

This thesis has four objectives, which will be researched sequentially.

- The first objective is to investigate what problems can occur in governmental healthcare systems that impede GDPR compliance in a crisis situation.
- The second objective is to determine the reasons problems in governmental healthcare systems occur according to the governmental personnel involved with these systems
- The third objective is to determine the specifications and requirements of an architecture that can be used in governmental healthcare in a future crisis situation, in relation to the GDPR
- The fourth objective is to develop a FAIR-based architecture that can be used to develop governmental healthcare systems in a crisis situation

1.5 Research Questions

Based on these objectives, the main research question for this thesis is: 'To what extent could FAIR guidelines improve governmental healthcare systems' compliance with the GDPR in a crisis situation?'. Which will be answered through the following sub-questions:

1. Which problems can occur that impede GDPR compliance in governmental healthcare systems in a crisis situation?
2. What are the reasons problems in governmental healthcare systems occur according to governmental personnel involved with these systems?
3. What are the specifications and requirements of an architecture that can be used in governmental healthcare in a future crisis situation, in relation to the GDPR?
4. How can a FAIR-based architecture for IT development of governmental healthcare systems in a crisis situation be developed?

1.6 Location

This thesis is a combined document for the Master of Science program 'ICT in Business and the Public Sector', and the Master of Science program 'Computer Science: Data Science', both given at Leiden University, in the Netherlands. The case study in this thesis and the produced framework are focused on the Dutch Healthcare sector, specifically the handling of the Covid-19 pandemic in the Netherlands. Sections 5 & 6 primarily cover the requirements for the first master, while Sections 7 & 8 cover the requirements for the second master.

1.7 Societal Relevance

The core of the social relevance of this thesis is that personal data, but especially healthcare-related data should remain both private and secure. During the Covid-19 pandemic, the GGD utilized systems that contain data about more than 10 million individuals, with tens of thousands of people processing information from this system. Given the scale of these systems, it is critical that actions are taken to ensure that the data contained in these systems remains private and secure as any breach could result in a negative impact on a very large group of people.

The systems used in the Covid-19 pandemic were already determined to not meet these standards, violating the GDPR. Data leaks have caused a lack of trust in the government's security practices. Consequently, fewer people are willing to interact with any such governmental system, which results in harm to public health. There is also a significant monetary impact, with the organizations managing these systems being at risk of significant fines from the data protection authorities [20] and a large class-action lawsuit against them [21].

This thesis aims to developed a FAIR-based framework to better comply with the GDPR, which would increase the probability that these standards are met. The other results of this thesis, especially the analysis of what went wrong with these ICT systems in Section 5, can also be used to prevent these same mistakes from occurring in the future. This can be applied to the corona systems that remain in use until the Covid-19 pandemic fully ends, as well as in any other future crisis situation. While this framework is designed to operate in a crisis situation, it could also be applied to non-crisis situations or other sectors, with the important consideration that the healthcare sector processes more sensitive information, which requires a level of security that may not be required elsewhere.

1.8 Academic Relevance

On an academic level, this thesis will be one of the first pieces of literature written about practically applying the GDPR to governmental healthcare systems. While some literature exists, there does not appear to be a practical implementation or framework to ensure compliance. The creation of a FAIR-based architectural framework to achieve GDPR compliance is also a topic little has been written about. Given that every system processing the data of EU citizens needs to be GDPR compliant, a scientific standard such as FAIR needs to operate in a GDPR-compliant way, which this framework could aid with. The created framework could also be used as a starting point for others to expand on. With future legal requirements being able to be incorporated into this same framework.

1.9 Ethical Considerations

This thesis will make extensive use of interviews to gather information that cannot be found in governmental documentation. For each interview, the interviewee will be notified of the conditions of the interview. The interviewee can stop the interview at any moment and for any reason and data will be processed anonymously and will not be able to be traced back to a specific individual, unless consent has been specifically provided. However, even if anonymized it's important to discuss how the interviewee will be cited as their area of expertise can be an identifier if the area is small.

As this case study involves a recent and major failing of the Dutch government, a certain level of anonymity must be ensured, so that individuals will not be afraid to share any information and that any statement made will not reflect on their career. After an interview has been processed, the interviewee will be contacted to make sure they are satisfied with the way they are referred to and how their words have been represented. And after this thesis is finished, each interviewee will be sent a copy when requested.

It is important to note that while individuals involved with the Dutch government were interviewed, at no point during the writing of this thesis was there any kind of governmental involvement. Nor was anyone paid for their participation in the interviews. The information stated in this report comes directly from official documents and the words stated by interviewees. It is however possible that the information from these official documents or interviews does not fully accurately reflect the truth of the situation. Given the amount of information gathered during this thesis, including the number of different sources, this effect will most likely be minor, if it is even present at all. Although it also cannot be excluded, without a governmental investigation with a higher level of access.

1.10 Research Design

This thesis will be in the form of a case study. The case study I have chosen is related to the GGD's handling of the Covid-19 pandemic in the Netherlands, through three different ICT systems (HPZone, HPZone Lite, and CoronIT) used to support the processes surrounding the Covid-19 virus. With HPZone Lite being a derivative of HPZone, sharing the same technical base. The case itself will be introduced in the next subsection. The GGD organization itself will be introduced in Section 4.4.

This case was chosen because this is the largest example of a governmental healthcare system in the location of this thesis, the Netherlands, which also has significant indications that the GPDR has been violated in numerous ways. As indicated by smaller-in-scope investigations by the Dutch government [22] [23], the supervisory GDPR organization 'Autoriteit Persoonsgegevens' in the Netherlands [24], a class action lawsuit against the organization that uses the system [25] and a previous investigation by me [26]. The Covid-19 pandemic is also the biggest and most recent example of the Dutch healthcare system facing a crisis situation, with these systems being used as ICT support systems.

The result of this thesis, while being about the GDPR, will focus on the Netherlands instead of the broader EU. The FAIR-based architectural framework will be based on the experience of Dutch governmental personnel. To make a usable architecture for use by the government in a crisis situation, it is important to take into account other regulations that also have to be met. As the GDPR is only a minimal requirement, with member states being free to create additional laws in this area, it is infeasible or even impossible to create an architecture that can be used EU-wide as-is, which means the focus will be on the additional Dutch regulations. Although other countries could likely benefit from at least part of the findings from this investigation.

1.10.1 Case introduction

In 2020, the Netherlands was confronted with a public health crisis. Covid-19, a highly infectious virus, spread through the Netherlands at a very rapid rate. The situation deteriorated significantly in the span of a few months - from worrying signals in January, to the first verified case in the Netherlands in February, to an 'intelligent lockdown' from March to April. The Dutch government was not prepared for such a situation, especially given the speed at which the Covid-19 virus spread and the severity of the disease itself, and responded by expanding the testing policy to national scale testing, in combination with source and contact tracing investigations.

This was supported through the use of three ICT systems maintained by the GGD. The first system, CoronIT, was acquired at the start of the pandemic to schedule appointments, register test results, and at a later stage of the pandemic register the vaccination status of the individual. The second system, HPZone, was already in use by the GGDs and it was used to conduct source and contact tracing investigations. The third system, HPZone Lite is a derivative of the previous system, limited to entries about Covid-19. These last two systems were only heavily used in the initial stage of the pandemic. A more detailed description of these systems can be found in Section 5.1.

Setting up and using these systems became a priority as the pandemic was continuing to get worse, with these systems giving the government a better chance to limit the spread. As the Netherlands had no experience with pandemics on this scale – the last global pandemic at this scale being the Spanish flue in 1918 and the last pandemic being the Mexican flue in 2009 which was limited in scale – there was no knowledge about how to best set up these systems. While source and contact tracing investigations were an already existing task, supported by an already existing ICT support system, the Covid-19 pandemic required this to be done on a significantly larger scale. Additionally, there was significant political pressure to develop and expand these systems as quickly as possible. Leading to many interruptions of service, crashes, and personal data breaches.

A significant data leak was reported in January 2021 by 'RTL Nieuws'. After months of internal warnings about the risks in the system had been ignored, it was proven that data from every single person in the system was at risk of being leaked. As well as that this had been happening on a large scale. Leading to the data potentially millions of people getting into the hands of criminals.

1.10.2 Proposed deliverables

Each of the four areas or research objectives has a distinct research deliverable, resulting in the following four deliverables:

- The first deliverable describes problems in areas of the GGD corona system that impede compliance with the GDPR. Based on the information gathered from investigating what exactly went wrong with these two Corona systems, including data access protocols, workflows, design choices, and information gathered from an internal interview.
- The second deliverable describes reasons why GDPR compliance is lacking and which areas need improvement. Based on the information gathered from two separate governmental conventions.
- The third deliverable is a set of specifications and requirements, based on what a healthcare system needs to contain to address the problems identified in the previous two research questions and two interviews with governmental officials involved in the healthcare sector
- The fourth deliverable is a FAIR-based architecture developed based on the previously identified specifications and requirements in addition to core FAIR concepts.

1.11 How can the generated knowledge be used?

This thesis presents a case study on the compliance of GGD corona systems with the General Data Protection Regulation. The findings of this study will be directly beneficial for the ongoing implementation of these systems, as they have been deemed unsafe by the Dutch Data Protection Authority and are still in use by the government. Although HPZone (Lite) is being replaced by GGD Contact in many locations and has been fully replaced in others already by GGD Contact, the improvements suggested in this study could potentially be applied to the successor system. Improvements could also be applied to CoronIT, which remains in active use. Furthermore, the FAIR-based architecture proposed in this study is highly customizable and can be applied to various crisis situations, making it relevant for future use as well.

The literature review has shown that smaller-scale healthcare organizations face significant challenges in fully complying with GDPR requirements. The FAIR-based architecture proposed in this study aligns with the principles of data minimization and privacy by design and by default, and can therefore help improve compliance with GDPR. There are few to no existing frameworks that can be used to improve GDPR compliance in healthcare systems right now.

The significance of this study is highlighted by the recent lawsuit filed against GGD by Stichting ICAM, which represents individuals who have interacted with the GGD system and whose data has been leaked [25]. The damages demanded amount to 3 billion euros, which far exceeds the budget for the system or any system that could have met GDPR requirements. In addition to this, the occurrence of personal data breaches has made a group of people distrustful of the government, preventing this group from participating in the testing and vaccination program. Implementing improvements suggested in this study could potentially mitigate this situation in the future.

2. Conceptual Framework

The conceptual framework for this thesis consists of the GDPR and the FAIR guiding principles for scientific data management and stewardship. The GDPR is the currently acting Data Protection Act in the EU and the evaluation criteria against which the GGD case will be evaluated. The FAIR guiding principles are principles designed to enhance the reusability of data holdings by making it easier for individuals and machines to find and use data and to support data reuse. However, as the use of these principles do not necessarily require an academic environment, this thesis aims to apply them in a broader context, resulting in the creation of a FAIR-based architectural framework that aims to address issues identified in the GGD case study.

2.1 GDPR

Before the introduction of the GDPR, data protection legislation in the European Union varied widely between member states. Data protection was specified in the 1995 Data Protection Directive, which specified the desired outcome without dictating the means and steps needed to achieve it. This variance led to "organizations doing business across the region [facing] a legal minefield of differing interpretations of data protection" [27, p. 5]. Technology has also changed significantly since 1995 when only 1% of the population of the European Union used the Internet [28].

The GDPR, which came into effect in 2018, aimed to standardize data protection legislation across all member states and expand the definition of personal data to include information such as IP addresses and internet cookies that did not exist or were not considered in 1995. Citizens also gained additional rights relating to their data, such as the right to request all information an organization has on them in an easily readable format, the right to request the transfer of this data to another company, and the right to request the deletion of their data, which will be expanded on in Section 2.1.1. Breach notification has become mandatory and standardized across all member states, with breaches needing to be reported within 72 hours or face fines of up to 2% of worldwide revenue or 10 million euros for minor breaches and 4% of worldwide revenue or 20 million euros for major violations.

In addition, the processing of data must now be closely monitored in each organization, supervised by the supervisory organization in each member state. Organizations with substantial data processing activities or those handling sensitive data must appoint an independent data protection officer who operates outside the company and ensures that all processing is directly necessary for company activities and that this information is communicated to the data subject. When processing sensitive personal data with a high risk of processing, a Data Protection Impact Assessment (hence 'DPIA') must be created before processing begins. This assessment describes the safeguards and security measures taken when processing data, the reason for processing, why there are no alternatives, and what additional measures are taken to address the risk and ensure compliance with the GDPR. If the independent data protection officer deems there to be a high risk that cannot be mitigated, the DPIA must be submitted to the supervisory organization for advice and/or judgment. The supervisory organization can also require the organization to make changes to its processing before they are allowed to start.

At the core of the GDPR are the principles of "privacy by default and by design" and "data minimization". Throughout the design, development, and processing of systems, as well as the composition of organizations, privacy must be considered and respected, and data processing must be limited to what is directly relevant and necessary for the system. It must be determined if the processing of personal information is even necessary, or if it can be processed using completely anonymized personal information. When personal data is required to be processed, it must be considered if some form of pseudonymization can be applied, in combination with strict access control and security measures. The development of systems must follow these principles and focus on security, and the processing of information must be organized safely and securely, including measures to identify aberrant behavior and detect data breaches.

The GDPR applies to any company or entity that processes personal data as part of their activities located in the EU, as well as any organization outside the EU that processes the personal data of EU data subjects. In practice, it can be difficult for organizations to separate EU citizens from other individuals, leading them to either block EU citizens from their services entirely or treat all individuals to the standards of the GDPR, the latter of which increases personal data protection for everyone. The adoption of the GDPR in the EU has also encouraged other countries and regions around the world to introduce their own data protection laws. The EU has been at the forefront of technological innovation and regulation in this area [29], with some major companies even attempting to make data protection a competitive advantage by changing their strategies to become leaders in data security products and services [30].

2.1.1 Data Ownership

Another important aspect of the GDPR is that it gives data subjects certain rights in the context of their data. While data may be processed at a location the data subject has no access to, the personal data that is being processed remains in their ownership. The data subject will always be informed about which personal data is being collected and processed, and for what purpose. In most instances, the processing of personal data may only occur when the data subject gives explicit permission. When processing personal data, the data subject has additional rights related to this data, the most important ones for this thesis being the right of access by the data subject (GDPR Article 15), the right to rectification (GDPR Article 16), the right to erasure (GDPR Article 17), the right to restrict processing (GDPR Article 18) and the right to data portability (GDPR Article 20).

While some exemptions exist, and they are clearly defined, organizations are required to defend their reasoning and should still strive to comply with these rights, wherever possible. In the context of this thesis, the following three exceptions are the most important, although even these exceptions follow rules, such as recital 156 [31, p. 29], that require appropriate safeguards for the rights and freedoms of the data subject: These exemptions are compliance with a legal obligation, the performance of a task carried out in the public interest or the exercise of official authority, reasons of public interest in the area of public health and archiving purposes in the public interest, scientific or historical research purposes or statistical purposes [31, p. 44].

Based on the right of access, at any point in time, the data subject can request any organization that processes their personal information what the purpose of the processing is, what the categories of personal data concerned are, to which entities personal data has been disclosed, the duration for which the data needs to be stored and the source of the information [31, p. 43]. The organization is also required to provide all personal data that has been collected and processed to the data subject.

Based on the right to rectification, at any point in time, the data subject can request any organization that processes their personal information to rectify inaccurate personal data concerning this individual. Including the ability to add missing information, through a supplementary statement [31, p. 43]. This right relies on the previous right as without this right it would not be possible to determine if any information is either incorrect or missing.

Based on the right to erasure, at any point in time, the data subject can request any organization that processes their personal information to erase personal data concerning this individual [31, pp. 43-44]. Unless any of the three previous exemptions would apply, with the appropriate safeguards for the rights and freedoms of the data subject being in place.

Based on the right to restriction of processing, at any point in time, the data subject can request any organization that processes their personal information to restrict the processing of personal data concerning this individual. This right applies when either the accuracy of the data is in question, the processing is unlawful and the data subject does not want to make use of the previous right to erasure, the controller no longer needs the data for processing but the data subject requires it for legal claims or the data subject has objected to the processing of their information, upon which it needs to be determined if this is a legitimate complaint.

Based on the right to data portability, at any point in time, the data subject can request any organization that processes their personal information to provide all personal data related to the data subject in a structured, commonly used, and machine-readable format. In addition to the right to transmit this data to any other controller. Although this right can only be utilized if the right to erasure could be used, as it requires the erasure of personal data related to the data subject at the controller.

2.1.2 Evaluation and Limitations

The European Commission published its first evaluation and review of the GDPR in June 2020, two years after its implementation [32]. Per Article 97 of the GDPR, such evaluations are required to be published every four years. However, given that the approval date of the GDPR is the 14th of April 2016, it appears that the EU chose this date as the start of this process.

The evaluation found that the GDPR has effectively strengthened individuals' right to personal data protection and ensured the free flow of personal data within the European Union [32, p. 4]. However, the Commission also identified areas for improvement, particularly concerning international transfers and cooperation between member states. It is currently too early to determine the overall effectiveness of the GDPR, and future evaluations will benefit from a longer period of application and experience.

The evaluation also specifically analyzed the effects of the COVID-19 pandemic on the GDPR. The GDPR permits certain limitations on personal data for public health purposes, but these measures must always be proportionate and uphold the fundamental rights and freedoms of individuals. The European Commission's conclusion regarding the GDPR concerning the COVID-19 pandemic is that, in the EU, the GDPR, in combination with the ePrivacy Directive [33], has been effective in providing practical solutions while maintaining a high level of protection for personal data.

The GDPR generally promotes competition and innovation by establishing a fair playing field for companies outside the EU. However, the implementation remains challenging, particularly for small and medium-sized enterprises. The proposal to create a compliance toolbox to help demonstrate compliance aligns with the aim of this thesis to develop an architecture that can assist with this issue. According to EU definitions [34], GGD GHOR Nederland, the executing party in this case, would be classified as a medium-sized enterprise, as it had 125 employees and an annual budget of 30 million euros at the start of the year when the project was signed in 2020, which is below the threshold of 250 employees and an annual budget of 50 million euros for a medium-sized enterprise [35]. However, during the COVID-19 pandemic, their budget increased significantly to 296 million euros per year and their headcount increased to 375 employees, which would classify them as a large enterprise. A change of this magnitude could lead to organizational difficulties.

The European Commission consistently emphasizes the need for Member States of the European Union to provide their national data protection authorities with adequate human, financial, and technical resources. Data protection authorities play a critical role in ensuring GDPR enforcement in each Member State, including cooperation between states in cross-border cases.

Although it has been reported that the Netherlands is one of the countries whose data protection authority, the AP, has seen the greatest relative increase in staff, this amount is still insufficient based on the organization's findings. In 2019, the AP was the third-largest data protection authority, with an annual budget of 18.6 million euros [36, p. 13]. A study commissioned by the Dutch Ministry of Justice and Security concluded that the AP has far too few resources to fulfill its duties, suggesting a budget increase to 66 million euros per year. As the Dutch data protection authority is already one of the largest in the European Union, a lack of funding here may indicate that this problem is also present in most, if not all, member states [37].

2.1.3 Scientific Research

Article 89 of the GDPR states that, provided appropriate safeguards are in place, data may be used and shared for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes. This allows for the use of personal data beyond the specific scope for which it was collected in the context of scientific research. However, organizations are still generally hesitant to share their data, and do not make use of this exception.

A recent example of this in the medical field is an investigation into excess mortality due to the COVID-19 pandemic. Based on a motion from Omtzigt on December 1, 2021, passed with unanimous support, this was to be investigated through an academic inquiry [38].

The RIVM and the CBS released a report on excess mortality during the COVID-19 pandemic in the summer of 2022. However, independent academic parties were unable to conduct their own investigation because they were denied access to government data, making it impossible to verify the results of the government organizations, discover new insights, and learn from the experience gained during the first two years of the pandemic.

This investigation is further complicated by the need to combine data from various sources, such as vaccination data from the RIVM and GGD test results from the GGD [39, p. 4]. Organizations often work independently, operating in silos that legally only allow for the sharing of data in a limited form unless permission is granted. According to a data scientist from the Ministry of Justice, data sharing can be done through a trusted third party (TTP), although not all organizations use one. In some instances, such as this investigation, the CBS fulfills that role. However, the CBS is not permitted to share this data for other purposes, such as investigations by academic parties [40].

The hesitation to share data for academic research is due to the RIVM's claim that, per the GDPR, data subjects have only permitted to share their data with the organization that collects it. GGD GHOR Nederland argues that their data can only be used by GGDs and GGD GHOR Nederland and that GGD GHOR Nederland does not have the authority to share data from individual GGDs [39, p. 6]. The GGDs themselves also considered the privacy risks of sharing to be too high [40]. While this is in line with the GDPR principle that data can only be used for the purpose for which it was collected, with the permission of the data subject, the exact enforcement of this varies between Member States [41]. It also fails to make use of the exceptions specified in Article 89.

In the Netherlands, the UAVG and the WGBO allow the use of data for research without explicit and specific permission from the data subject. However, it is unclear which data can be shared under which circumstances, for what purposes, by whom, and which technical and organizational measures must be taken [39, p. 7]. What is clear is that there is a need for better data access [42] and that not utilizing data can come at a significant cost. For example, the lack of data sharing in the UK is estimated to contribute to the deaths of thousands of people and billions of pounds in financial burdens to society [43]. With the Netherlands likely exhibiting similar, although smaller, effects.

While access to data is a major problem for the reuse of data in academic research, other challenges must be addressed as well. Even with the exceptions discussed earlier, explicit and specific permission from the data subject remains a significant obstacle. Other issues include the quality and completeness of the data, the lack of interoperability of data, and the resulting costs of resolving these issues, all of which make it difficult to reuse healthcare data for academic research [39, p. 7].

2.2 FAIR

The FAIR Guiding Principles for scientific data management and stewardship provide a comprehensive, high-level framework for organizing, managing, and sharing data. Outlining the fundamental principles that should be followed to ensure responsible, consistent, and effective use of data resources. The principles emphasize the need for data to be Findable, Accessible, Interoperable, and Reusable (FAIR), and they provide a basis for organizations and individuals to build sustainable, effective data management practices [44].

The FAIR guiding principles are a relatively new standard, as they were created in 2016. However, FAIR has already resulted in numerous collaborations as well as extensions of the original framework. One of them is FAIRDOM, which is a platform that is built on the FAIR principles and provides a suite of tools to help researchers share, manage, and analyze their data [45]. Based on the FAIRDOM-SEEK platform, there are various platforms such as FAIRDOMHub [45], DataHub [46], and others [47] [48] [49] [50] that operate as a repository for FAIR-based datasets used in publications.

The FAIR guiding principles are also applied in a less academic environment. One such example is the Virus Outbreak Data Network (hence ‘VODAN’) project that addresses various issues that are common to health data management across the African continent. By adopting the FAIR principles “clinical and research data that has been locally generated, curated and held on-site at health facilities, partners or respective ministries can be made available to the rest of the world” [51]. As this project was developed to aid in fighting the Covid-19 pandemic and improves data access, in addition to being used for medical decision-making, the application of the FAIR principles may also be able to improve the handling of medical data in the Netherlands.

Based on the success of this project and others, the Dutch government is already investigating if the application of FAIR principles can improve data handling in the government in general. For example, ZonMW, the Dutch organization for health research and healthcare innovation, is studying FAIR as a possible improvement to the data access problem explained in the previous section [39, p. 2]. Forum Standaardisatie, which is a Dutch governmental organization that researches and establishes security standards, has also investigated the use of FAIR in the government as a whole and concluded that “It is not a question of if organizations have to work with the principles, it is a question of when and in what way?”, adding that a concrete elaboration of the principles is now required [52, p. 28].

This thesis intends to apply these guidelines, including the additional concepts such as "federated data", "data visiting" and "machine accountability, to not just the academic aspect of data management, but to apply them in an architectural framework that can be used by organizations. The FAIR guiding principles, as proven by these examples, don't necessarily require their use in just an academic setting. Therefore, a more concrete architectural framework might have significant importance and value in areas processing sensitive data such as the healthcare sector.

2.2.1 Principles

The principle of Findability is focused on ensuring that digital resources can be found and identified quickly and easily. Which applies to both humans and machines. The various sub-principles that make up this principle have been listed below.

- Principle F1. (meta)data are assigned a globally unique and persistent identifier
- Principle F2. data are described with rich metadata (defined by R1 below)
- Principle F3. metadata clearly and explicitly include the identifier of the data it describes
- Principle F4. (meta)data are registered or indexed in a searchable resource

Based on these four sub-principles, digital resources need to be indexed by search engines and other discovery tools using a unique identifier, they need to be described with detailed metadata and they need to be properly linked and labeled. Which must be done in a manner that respects the principle of Reusability.

This principle can also be extended to the design of the application or website itself. A powerful search engine, that is complicated for its users, will not be able to find and identify digital resources quickly and easily and as a result, would not fully adhere to the principle of Findability. Therefore, to ensure this principle, users of these systems need to be provided with clear and easy-to-understand navigation methods and menus, as well as use descriptive and meaningful tiles, labels, and keywords.

The principle of accessibility is focused on making sure that digital resources are accessible to users, either human or machine. The various sub-principles that make up this principle have been listed below.

- Principle A1. (meta)data are retrievable by their identifier using a standardized communications protocol
    - Principle A1.1 the protocol is open, free, and universally implementable
    - Principle A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- Principle A2. metadata are accessible, even when the data are no longer available

Based on these sub-principles, the digital resources that were identified based on the previous principle need to be made accessible to the user. While these digital resources must be easy to access, it does not mean that everyone should have access. Due to the highly valuable nature of healthcare data and the danger to data subjects, if this data were to be openly shared, the storage and access of this data must be done with privacy and security in mind, limiting access to users that are both authorized AND need the data in the context of their job.

This must be achieved through a properly designed security system, using authentication and authorization procedures. At a minimum, these authentication measures must include both a password and a two-factor authentication method. The data itself must be encrypted to ensure that data is only accessible via this security system. On top of this, data access and functionality must be designed with the role of the user in mind. This is achieved through a role-based access scheme that can give different levels of access to different users.

To verify all this and ensure the integrity of the data, user activity must be logged and analyzed. By following these principles, automated systems can both detect suspicious activity and act upon it, with the resulting audit trails making it possible to investigate activity/users when required.

Based on the final sub-principle, systems should include measures to ensure the availability of data, which can be achieved through redundant storage and extensive backup systems. While ensuring that the backup system complies with the GDPR. Through these methods, the chance of data becoming inaccessible due to systems becoming unavailable is significantly reduced. Data should also be processed in a way to benefit from the data while the raw data is not accessible anymore, as keeping healthcare data for a longer time increases the risk to privacy

The principle of Interoperability is focused on ensuring that digital resources can be exchanged and used across different systems and applications. Allowing two or more systems, applications, or services to communicate, exchange data and use the data that has been exchanged. A common standard relating to how data is stored and defined allows for data sharing between different organizations and stakeholders.

- Principle I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- Principle I2. (meta)data use vocabularies that follow FAIR principles
- Principle I3. (meta)data include qualified references to other (meta)data

Based on these sub-principles, systems should ensure that they are using a shared vocabulary so that the information within systems can be shared without the need for data transformation and other methods to convert the data into the required format. Which needs to be an established standard that all organizations are or will start using in the future.

In the context of Dutch healthcare data, interoperability is essential for ensuring that the data is secure and confidential. To ensure that the data is secure and confidential, the standards for interoperability must comply with Dutch data protection laws. The standards for interoperability must also take into account the responsible use and sharing of data. Full interoperability is difficult to achieve as it requires not just the current system to adhere to the FAIR principles, but all others as well. Otherwise, data is still not interoperable.

The principle of Reusability is focused on enabling organizations to reuse digital resources, processes, services, and components to generate new knowledge or apply this data in service of creating new applications and services. Improving the ease or reusing of data, encourages collaboration and reduces duplication of efforts. The various sub-principles that make up this principle have been listed below.

- Principle R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - Principle R1.1. (meta)data are released with a clear and accessible data usage license
  - Principle R1.2. (meta)data are associated with detailed provenance
  - Principle R1.3. (meta)data meet domain-relevant community standards

For the reuse of data, anyone interested in using this data must have a clear description of the data to ensure that the data is relevant to their task, that the data is allowed to be reused, and under what conditions the data is allowed to be reused as well as who to credit is data is ultimately used to support any process or investigation [53, p. 14]. From the example of the unwillingness to share data with the academic world, the reusability aspect should focus on the specification of the conditions under which reuse is acceptable and the form in which data can be shared without endangering the rights of the data subject. This can be supported through the use of techniques such as ‘Data visiting’ and ‘Federated Data’, which will be introduced later in this Section.

2.2.2 Data Federation

Using FAIR, organizations can benefit from data from across their entire organization, while not having to transfer their data to central locations. This is achieved through a federated approach to data management, called data federation, that combines autonomous data stores to form one large data store. Based on the concepts of ‘data virtualization’, ‘Heterogeneous set of data stores’, ‘Autonomous data stores’, ‘One integrated data store’, and ‘On-demand data integration’ [54].

The concept of data virtualization refers to a data management approach where data can be retrieved and manipulated without requiring information about the data, specific formatting, or the location where it is physically stored. Data is accessed in real-time but is not moved or changed in any way. This is in contrast to a traditional Extraction, transformation, and Load approach, where data needs to be extracted from their database, transformed, and cleaned into a common format that removes irrelevant entries, and then loaded into a database. As no data is moved or duplicated, this significantly reduces the potential for data duplication errors. And as not all data needs to be stored at a singular location, this reduces the impact of a personal data breach if it were to occur.

The concept of a Heterogeneous set of data stores refers to the concept that this approach should make it possible to combine data from various data stores using different storage structures, different access languages, and different APIs. Users should be able to access different types of database servers and various file formats.

The concept of Autonomous data stores refers to the fact that while different data stores need to be able to form one large data store, it remains important that they can operate independently. If data stores can’t operate independently anymore, while data may not be physically located in the same location, a centralized database has been constructed instead of a federated one. This is important in the healthcare sector as clinics, hospitals, and other healthcare locations need to remain independent entities. Only when data needs to be accessed by a different entity should the data federation system approach be (required to be) used.

The concept of one integrated data store refers to the concept that no matter where or in what format data is stored, it should be returned to the user as a single integrated data set. As data could have been stored in various formats and methods, data federation requires data to be transformed into a common format, so that this data can be used. And as long as this step is taken at the end, the actual storage of data does not have to align with this common format. However, to ensure the FAIR principle of interoperability, the focus should be on creating and applying common standards between all data stores.

The concept of on-demand data integration refers to the fact that when a user requests data, this data is not combined beforehand in any kind of system, but is instead retrieved and combined at that moment. And when this data is requested, it does not modify the original data in any way, which remains at its original location, in its original format.

This data management approach can also be extended to not just giving access to the data itself, but to limit access to only the results of each data query and combine this in a single result. For instance, when information about the number of items matching a query is required, or when a distribution of data needs to be created, it isn't necessarily required to see or have access to this data. With data federation, the results from each data repository could be shared and then combined into a single source. Which achieves the same result as having access to all data, while maximizing the privacy and security aspect.

In the context of this case study and the healthcare sector in general, this data approach is particularly interesting. The GGD already consists of many individual organizations, each doing a similar job with results that need to be combined between all of them. As this structure matches the separation of health facilities in different countries in the VODAN AFRICA project, where data federation is already being applied [55], data federation might be able to be directly applied to the GGD too.

2.2.3 Data Visiting – Personal Health Train

In traditional healthcare-related scientific research, when data from different sources and/or from multiple countries is required, individual registries would be asked to share their data which would then be brought together in a central repository. However, this method results in "a lot of double work in maintaining the various resources", "synchronization issues with resources that were already there" and "legal constraints" that inhibit the sharing of data across national borders.

The FAIR guidelines, in combination with the previous aspect, address this problem using a data-visiting approach. In a data-visiting approach, applications containing questions and algorithms are sent to the data source, instead of data being sent to the user [56]. A currently used data visiting approach in the GO FAIR organization is the Personal Health Train (hence 'PHT') [57], designed to improve the use and reuse of health data by enabling healthcare professionals and scientific researchers to be able to work with health data from various data sets. Which can also "give controlled access to data, while ensuring privacy protection and optimal engagement of individual patients and citizens" [58]. In addition to this, the VODAN Project also makes use of this approach for medical decision-making as a non-scientific activity.

The rest of this section will discuss numerous benefits related to the data visiting approach, especially regarding controlling access to data and the ability to display data in a more secure form compared to the data being processed. These are in addition to the benefits stated by the creators of this approach, such as avoiding double data entries and ambiguity, and through the principle of interoperability, standardization of the dataset already stated in its design [58].

In contrast to an approach where data files are being shared, which would lead to the organization losing control over their data, an organization that uses the data visiting approach will always have full control over how much of their data is being made accessible to other entities. It also gives the possibility to revoke this access in the future, something which can be negotiated by contract with traditional data sharing but leaves the organization with no guarantee that this data will never be accessed again. With the combination of a logging system, a Dutch requirement under the NEN 7513 when processing medical files, this approach ensures that it is possible to keep a log of every single interaction with the data, which would have been impossible when individual files would be shared.

Another advantage of this approach is that while data files that are being shared can be outdated, a real-time data connection with the actual data source cannot. Using the data visiting approach it can always be assured that all information is the most up-to-date, which ensures that errors are corrected when they are detected and that the latest entries are included. It can even allow for most

GDPR data ownership rights to be followed as instead of needing to change already shared data files, changes can now easily be made in real-time.

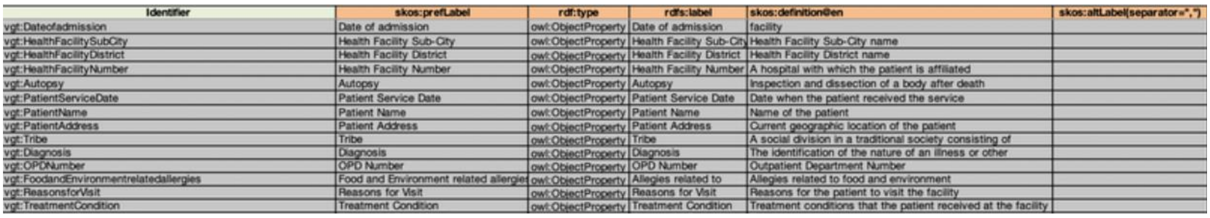
Another advantage of this approach is the fact that it allows for data shown to the user, and data used in the calculations to be different from each other. This can be used when there is an additional need for privacy and/or security, where it would not be allowed to share certain data, but it would be allowed to share a transformed version of the data. Following this process allows for academic research on data that would normally never be accessible to researchers due to various reasons. And all of this without compromising the privacy or security of the data subject.

2.2.4 Machine Accountability and Metadata

The previous elements, and FAIR in general, rely on well-structured metadata that covers not only what type of data is stored but also describes when and for what purpose data may be accessed, and most importantly it connects to, ideally, an already existing vocabulary that can be used to combine data from various sources, even if the exact format varies among these sources. Without this metadata structure, a FAIR architecture would not function.

Based on the first principle of Interoperability, all digital resources need to be accessible using a common language. The exact language is not important, as long as it is a common one shared between all organizations. However, benefiting from an existing standard may be easier than creating a new one and convincing others to use one. The VODAN-Africa Project uses a mature standard called Simple Knowledge Organization System (hence 'SKOM'), to produce the vocabulary that references these digital resources, which is then connected to a CEDAR template that connects these digital resources to a formal definition for each entry [59, p. 4]. This metadata is represented in JSON-LD, as well as Resource Description Framework (hence 'RDF') triples [60].

An example of this can be seen in Figure 1 [61]. The identifier column contains the formalized definition that is shared between all organizations giving access to their data. The skos:prefLabel contains the internal label as it is mapped to the identifier. The rdf:type column contains the type of data the digital resource is. The rdf:label contains the label that will be displayed in the internal system. The skos:definion@en column provides an English definition of the digital resource. Using this structure, it becomes possible to ensure that organizations are talking about the same concept, independent of the structure that is stored in, the location, and the language that is used.



Identifier	skos:prefLabel	rdf:type	rdfs:label	skos:definition@en	skos:altLabel(separator=",")
vgt:Dateofadmission	Date of admission	owl:ObjectProperty	Date of admission	facility	
vgt:HealthFacilitySubCity	Health Facility Sub-City	owl:ObjectProperty	Health Facility Sub-City	Health Facility Sub-City name	
vgt:HealthFacilityDistrict	Health Facility District	owl:ObjectProperty	Health Facility District	Health Facility District name	
vgt:HealthFacilityNumber	Health Facility Number	owl:ObjectProperty	Health Facility Number	A hospital with which the patient is affiliated	
vgt:Autopsy	Autopsy	owl:ObjectProperty	Autopsy	Inspection and dissection of a body after death	
vgt:PatientServiceDate	Patient Service Date	owl:ObjectProperty	Patient Service Date	Date when the patient received the service	
vgt:PatientName	Patient Name	owl:ObjectProperty	Patient Name	Name of the patient	
vgt:PatientAddress	Patient Address	owl:ObjectProperty	Patient Address	Current geographic location of the patient	
vgt:Tribe	Tribe	owl:ObjectProperty	Tribe	A social division in a traditional society consisting of	
vgt:Diagnosis	Diagnosis	owl:ObjectProperty	Diagnosis	The identification of the nature of an illness or other	
vgt:OPDNumber	OPD Number	owl:ObjectProperty	OPD Number	Outpatient Department Number	
vgt:FoodandEnvironmentrelatedallergies	Food and Environment related allergie	owl:ObjectProperty	Allegies related to	Allegies related to food and environment	
vgt:ReasonsforVisit	Reasons for Visit	owl:ObjectProperty	Reasons for Visit	Reasons for the patient to visit the facility	
vgt:TreatmentCondition	Treatment Condition	owl:ObjectProperty	Treatment Condition	Treatment conditions that the patient received at the facility	

Figure 1 - Example of a Machine accountable Metadata structure

2.3 Overlap and Distinction between FAIR and the GDPR

As the GDPR is a legal requirement, the FAIR standard by definition needs to be GDPR compliant if European personal data is being processed. If this is not the case, then the FAIR standard may not be used at all, at the threat of significant fines. By comparing the concepts of both FAIR and the GDPR, this section will explain why FAIR has the potential of improving GDPR compliance, while it should still be noted that FAIR is not GDPR compliant by default. It relies on organizations applying FAIR in a GDPR-compliant way.

The principle of data minimalization can be achieved with FAIR, however, this is not by default nor is FAIR able to cover the entirety of the data processing processes to be able to ensure data minimalization. With FAIR, organizations can adhere to data minimalization by carefully selecting which data will be made accessible to other entities, or even to themselves, which complies with the principle of data minimalization. However, this relies on the organization following this process in a GDPR-compliant way, the technique itself doesn't ensure compliance, but a correct implementation of the technique does. As FAIR does not cover the way information is collected, only the way that data is stored and made accessible, it also can't ensure that only information that is required for a specific purpose will be collected.

In addition to only providing access to specific data, data minimalization is also achieved using the federated data concept. As data can be accessed in real-time, over multiple sources, it is no longer required to store duplicates of information in a centralized location. There is also no need to store the information in a different, more common form, in addition to the original data. This reduces the amount of data that is stored about a data subject, without impacting the use of the data. However, queries likely need to be stored to record what was done with the data, which would increase the amount of data stored about the data subject. Although to a much lesser extent than any kind of data sharing approach, where actual databases are sent to different locations.

The various data ownership-related rights that data subjects have under the GDPR can also be achieved with FAIR, serving as an improvement over the current situation. With a data visiting approach instead of a data sharing approach, organizations always remain in full control over their data. As a result, it becomes significantly easier to share all related data to the data subject, including who accessed the data and for which purpose, as they aren't required to consult other organizations anymore. It also becomes far easier to adjust an entry or delete it in its entirety or restrict forms of access as it does not require making changes to data files that were already shared. Changes are easily made at their organization, which then apply to anyone given access to their data. The interoperable structure of data would also make it easy to follow the right to data portability. The only possible problem with this approach is that it does not address the fact that organizations are unwilling to give access to their data for fear of liability or violating the GDPR, it only makes it safer to do so.

The principle of "privacy by design and by default" can also be achieved with FAIR, however, it depends on its implementation. The federated data approach reduces the number of locations where data is stored and would prevent the need for a centralized data store with a large amount of information, but it still requires each data store connected to the federated data network to take appropriate security measures. The data visiting approach eliminates the need to share data files to give access to data, but it does not prevent malicious actors from accessing information if the organization that received access to data is compromised themselves. And if more access to data is given then would be warranted, the system would still be regarded as incompatible with the principle of "privacy by design and by default", even though the design of the system can limit access to data.

3. Methodology

The methodology for this thesis has been separated based on the four sub-questions of the main research question. Sections 3.1 to 3.4 describe in detail how each of these sub-questions was answered, including how each data source was used to gather a certain type of information.

An overview of these sections has been included at the end of this Section, depicted in Table 1 to Table 4. Each table corresponds to a research objective, matched with the corresponding research question. Including the data type of the sources used for information. The final column evaluates each source on its reliability, representativity, and validity.

3.1 Problems that inhibit GDPR compliance in a crisis situation

To investigate the various problems that inhibit GDPR compliance in this case study into the corona systems of the GGD, the most valuable source would normally be the various DPIAs created for the various systems. However, these sources were not able to be properly used. The DPIA for CoronIT was not continuously updated throughout the development process, with the latest version being updated last somewhere between the start of 2020 and November 2020. While risks and mitigation methods were stated, including the estimated risk level for the data subject, these methods were not implemented yet at the time of the creation of this report. Which makes it difficult to evaluate the system as it was actually implemented. HPZone and HPZone Lite, the other two systems evaluated as part of this case study, have not had a DPIA created as part of their development process. There is a DPIA that was created at one GGD sub-organization after the data leak had occurred, however, this document already took into account certain improvements that were made and does not represent the state of the system as it was before the data leak. Therefore, other sources of information were required to be used.

For this research objective, it is more important to determine which problems have occurred in which areas instead of determining the exact number of violations. The data required to determine how many times an article has been violated will likely either not exist or will not be accessible, even to the government itself

This investigation consists of information gathered from a comprehensive media investigation covering reporting about the GGD and the various systems used during the Covid-19 pandemic, as well as governmental documentation about and governmental investigations into this matter, as well as correspondence with governmental personnel and originations and an in-depth interview with two individuals closely involved in this case when they were active on the management level of one of the GGDs during the early stages of the pandemic.

Multiple different sources of information were used to apply data triangulation, with each source having some kind of drawback. The media investigation is unlikely to cover all information, as media items have to be newsworthy and generally do not go into detail. Governmental documentation meanwhile is comprehensive but depends on the areas that have been investigated and is limited to what has been made accessible to the public or requested by WOO requests. The WOO requests I submitted are based on a lack of information in areas I am aware of, however, other useful information may exist in areas I wasn’t aware of. Governmental correspondence has been used to find information that was not directly publicly available but depends on what other information has been gathered to be able to ask detailed questions and it depends on the organization's willingness to answer. The in-depth interview offered valuable insight into the inner workings of the GGD, but this view in one of the GGDs may not apply to the other GGDs in other regions. However, all sources combined should give as clear a view of the systems used by the GGD as possible for any independent governmental investigation.

To aid this investigation, a timeline of events related to this case has been established. Which is listed in Section A4 of the appendix. The timelines consist of events related to the Covid-19 pandemic in general, as well as information related to the GGD, data breaches, the development of systems, and the governmental response to the various issues.

3.2 Governmental personnel response to the lack of GDPR compliance

To investigate this research question, I attended two governmental conferences related to ICT projects and Data management. Where I participated in various presentations and panels as well as conducted informal interviews of various lengths. These conferences offered the opportunity to approach many individuals involved in a high level of government in an informal setting, from a variety of different ministries and departments. Individuals in general were very approachable, only limited by the amount of time.

However, most interviewees have been cited anonymously. This allows all interviewees to speak out without fear of repercussions, as all interviewees are individuals currently active in various governmental organizations. An open interview could influence their willingness to answer questions truthfully and/or in full. Every interviewee was made aware of these conditions and based on this, some of them were willing to be cited by name. However, it cannot be determined if some individuals would still be inclined to not speak out for fear of repercussions.

The first governmental conference was iBestuur 2022, the largest convention of top-level governmental employees and organizations involved in the digitalization of the Dutch government. Topics discussed during this conference were among others, the digital transformation and the relationship between the government and society, and societal challenges related to IT and cybersecurity. Including talks from Hennie Brands, CIO of the Ministry of Justice and Security, Paul van Kruistem, CIO of the National Coordinator for Security and Counterterrorism, Alexandra van Huffelen, Secretary of State of Kingdom Relations and Digitalization, and Ron Roozendaal, the 2022 Deputy Director General on Digitalisation and former CIO of the Ministry of VWS. At this conference, I was also able to speak with the highest-ranking governmental individual in this area, the Secretary of State of digitalization, Alexandra van Huffelen.

The second conference was a conference organized by the Ministry of Justice and Security and primarily focused on the various organizations that fall under the umbrella of this same ministry. Informal interviews with the various organizations in attendance gave an overview of data security standards throughout this ministry. Of special note is the informal interview with the Chief Data Officer of the Ministry of Justice and Security, Ronald Damhof, which offered specific security measures related to information access by employees. I also participated in a panel led by Ronald Damhof, which connected a dozen professors, legal experts and leading governmental employees on the matter of data standards and academic research.

3.3 Specification and Requirements of the architecture

To determine the specification and requirements of what is required to successfully manage projects, the findings from the previous two research questions have been combined with two interviews. The first research question indicated various points of failure in the GGD case, which was taken into account for the new architecture model. The second research question indicated some more general issues in governmental project management and various methods that were already in use. This information was then used as a basis for the questions for the following interviews.

The first interview was an interview with Mustafa Kedilioglu who is a program manager in the Ministry of the Interior and Kingdom Relations. Of particular interest was his role as project manager for the Corona Melder App, discussed in Section 4.6, which was an investigation by the government to draw lessons from to be applied to governmental IT development in other projects. The goal of these interviews was to gain practical knowledge related to ICT development. Which includes determining the existing architecture that is already used by the government in a crisis. As well as to gain knowledge about project management in governmental projects and identify where improvements can be made. As the result of this thesis would be implemented by project managers, this information is of significant importance.

The second interview was part of an interactive day where I joined Ron Roozendaal during his activities as the 2022 Deputy Director General on Digitalisation. Both his current function and former function as CIO of the Ministry of VWS were very valuable for this thesis. In his position, as Deputy Director General on Digitalisation, he was involved with improving the digital aspects of the Dutch government, which is closely related to my thesis topic. In his former position as CIO of the Ministry of VWS, he was closely involved with the various systems developed and used at the GGD, including CoronIT, HPZone, HPZone Lite, GGD Connect, the Corona Check app, and the Corona Melder app. However, the development of at least CoronIT and HPZone (Lite) was led by GGD GHOR. Due to his high position in the Dutch government and close relation to the case, this interview had significant value. And it was also relevant for establishing the difference between the management of projects at a project manager level and the top level of an organization.

3.4 FAIR-based architecture

This architecture aims to create a solution based on the FAIR guiding principles that address the challenges faced by the Dutch government in managing healthcare data, in a crisis situation. Elements from the previous research question have been examined to determine if they could be improved by introducing concepts from the FAIR guiding principles. These principles have also been extended with concepts closely related to FAIR, as discussed in Section 2.2. The VODAN-AFRICA project has been used as an inspiration for this architecture, particularly in terms of data visitation and federated data management, as well as their accessibility scheme. This project is of particular interest to this architecture as it is a mature FAIR-based project, working with Covid-19 healthcare data.

To ensure that the architecture is usable for the Dutch government, it is important to consider the existing governmental architecture and initiatives related to this area. Additionally, the final solution should be highly customizable and not reliant on advanced technology, as it will need to be applied in crisis situations. A participatory approach has also been taken in the development of this architecture, involving the discussion of elements of FAIR and the FAIR-based architecture with various governmental employees, such as project managers, a data protection officer, and high-level government officials in the area of data management and FAIR experts. The most notable FAIR experts I have spoken to are my first supervisor Prof. Dr. Mirjam van Reisen based on her experience with FAIR Data Science and her involvement with the VODAN-Africa project and my second supervisor Dr. Katy Wolstencroft based on her experience with semantic data and knowledge integration for biomedical data science, with a particular focus on FAIR.

Research Objective	Research Question	Data Type	Data Sources	Data Reliability, Representativity, and Validity
Investigate what problems can occur in governmental healthcare systems that impede GDPR compliance in a crisis situation	Which problems can occur that impede GDPR compliance in governmental healthcare systems in a crisis situation?	Secondary – Public Records	Data Protection Impact Assessment (WOO)	• Reliability: Expected to be moderate-high ○ While the information from secondary sources should give an accurate view of the situation, it is important to note that these are only public records. It is possible, likely even, that information listed here is incomplete. ○ 'WOO' requests were used to request information to be made part of the public record, but for these requests, you need to know what to ask for. Which makes it possible that some additional information remains inaccessible. ○ As this is a situation that is currently still ongoing, it is possible that new articles will be released, resulting in slight differences depending on when an investigation was started. Articles might also be missed. However, this should not result in a significantly different outcome. • Representativity: Expected to be low-moderate. ○ There is a major problem in ICT in government in general, leading to widespread project failure, which will likely be present here. ○ While the results are specific to this single case, it may be possible to identify issues that other health data handling organizations may also have ○ As the pandemic is an unprecedented situation where the government had to act quickly, its possible mistakes were made due to the speed instead of a systemic issue ○ Even in the context of this case, the information gathered might only reflect the results of that specific region instead of the overall project. • Validity: Expected to be high. ○ As any violation of the GDPR is a result of an area that is not in compliance, validity in determining areas that aren't compliant is high ○ The combination of public information, media articles, and correspondence should be enough to get an accurate view of the situation. ○ The difficulty is in determining if areas are fully compliant with the GDPR. As it might only be a result of the information being incomplete or region-specific
			Project Tender	
			Documentation by parliament	
			Documentation by the data protection authority, the AP	
		Secondary – Other Material	Newspapers	
			Correspondence with supervisory organizations	
		Interviews – Semi-structured	Management-level personnel from the GGD	

Table 1 - Data Collection and Data Analysis Overview – Research Question 1

Research Objective	Research Question	Data Type	Data Sources	Data Reliability, Representativity, and Validity
Determine the reasons problems in governmental healthcare systems occur according to the governmental personnel involved with these systems	What are the reasons problems in governmental healthcare systems occur according to governmental personnel involved with these systems?	Interviews - Informal	iBestuur conference attendees	• Reliability: Expected to be High o Due to the number of people being interviewed, findings should be reliable and cover a significant extent of the problems in governmental systems • Representativity: Expected to be High o As these questions are related to the general problem, instead of being related to the case, answers should be representative of the general problem o The problems in the case may occur due to different reasons than the reasons for the general problem. • Validity: Expected to be moderate-high o While the interviews will be anonymous, it's still unlikely that all individuals will be entirely honest in interviews due to fear of repercussions o Interviewing a variety of individuals should indicate the dominant view
			Ministry of Justice and Security Data conference attendees	

Table 2 - Data Collection and Data Analysis Overview – Research Question 2

Research Objective	Research Question	Data Type	Data Sources	Data Reliability, Representativity, and Validity
Determine the specifications and requirements of an architecture that can be used in governmental healthcare in a future crisis situation, in relation to the GDPR	What are the specifications and requirements of an architecture that can be used in governmental healthcare in a future crisis situation, in relation to the GDPR?	Secondary– Public Records	Published Literature	• Reliability: Expected to be high o As interviewees are currently active in the field of project management, I expect that findings should accurately reflect what's needed to successfully develop projects • Representativity: Expected to be High o As the specifications and requirements are related not only to the case study but to the general situation as well, answers should be representative of the general situation • Validity: Expected to be Moderate-High o As this objective is not related to any project failing, it is unlikely that people will not be truthful with their answers. o However, there may be contrasting views on exactly which changes need to be made to successfully manage projects o Therefore, the specifications and requirements may vary from person to person, which should be mitigated by doing several interviews
		Interviews – Semi-structured	Governmental personnel experienced in Project management	

Table 3 - Data Collection and Data Analysis Overview – Research Question 3

Research Objective	Research Question	Data Type	Data Sources	Data Reliability, Representativity, and Validity
Develop a FAIR-based architecture that can be used to develop governmental healthcare systems in a crisis situation	How can a FAIR-based architecture for IT development of governmental healthcare systems in a crisis situation be developed?	Secondary–Public Records	Published Literature	• Reliability: Expected to be Medium o There is a lot of potential for different designs, although they might be equally valid o The interview(s) for this research question doesn't have to be anonymized. The same questions could be asked to the same people generating the same answers. However, differences will still occur due to information generated from where the discussion leads based on those questions • Representativity: Expected to be High o For both the creation of the architecture and to see if it complies with the GDPR, it's necessary to do a careful review of the literature. These are both limited by their respective age, resulting in less academic information published in these areas. However, general conclusions should be able to be made. o The architecture will relate to governmental healthcare systems in a crisis situation as a whole, instead of only the case study • Validity: Expected to be Moderate o There is currently no real usable framework with compliance with GDPR with current technology so there isn't much to compare to. o FAIR has only existed since 2016 o Given that the architectural framework based on FAIR is more generalized than the GGD case, interviews with experts should result in a model that satisfies the research objective
		Interviews – Participatory Design	Governmental personnel experienced in Project management	
			FAIR experts	

Table 4 - Data Collection and Data Analysis Overview – Research Question 4

4. Context

Governmental IT in the Netherlands has been a concern for the past decade and the implementation of the GDPR has added further complications, as highlighted by the Autoriteit Persoonsgegevens (hence 'AP') and the Bureau ICT Toetsing (hence 'BIT') / Adviescollege ICT (hence 'AcICT'). The GGD, a unique federated semi-governmental organization that combines regional organizations, with complete autonomy in their region, to operate on a national level, serves as an interesting example for the case study. The GDPR's requirements and regulations, including the Data Protection Impact Assessment, will be taken into account in the analysis of the GGD's IT systems.

Additionally, the corona systems utilized by the GGD are not one-of-a-kind. In a typical scenario, vaccination programs would be supported by Praeventis. Therefore, it is relevant to have an understanding of this system and why it was not selected for use in the Covid-19 pandemic. In response to the pandemic, another system called the Corona Melder App was developed, which did not exhibit flaws that will be discussed in this thesis. Rather, it was stated to serve as an example of how government IT projects should be developed. Although it should be noted that this system was primarily automated, with little to no human involvement in processing personal data.

4.1 State of governmental IT in the Netherlands

The Dutch government has faced major problems for the past decade due to a shortage of internal IT personnel and the resulting lack of in-house ICT knowledge. In the future, this issue is expected to become even more pronounced due to both the need to expand the government's IT workforce and the need to replace staff members who are retiring [62]. To further investigate this, the Dutch government conducted a parliamentary inquiry in 2014 to investigate the failures in governmental ICT-project development, led by the commission Ton Elias [63]. This inquiry aimed to identify common threads and patterns of errors to provide solutions and prevent such mistakes from occurring in the future. As the most recent in-depth investigation on this subject, it is regarded as an important benchmark by and for the Dutch government.

The overall ICT organization in the central government was found to be chaotic and opaque, with fragmented and unclear tasks and responsibilities. The interests of key players in ICT projects often diverged, and the national government lacked control over the cost, time, and outcome of these projects. The current culture surrounding ICT projects in the national government is also unsustainable. On the one hand, there is an over-enthusiasm for ICT, in which it is seen as a solution to all issues. On the other hand, the House of Representatives often calls for policies without considering the technical feasibility of their implementation. As a result, ministers often commit to implementing these policies without considering whether they are technically feasible.

This situation is made worse by the fact that the national government often ignored the expertise of ICT suppliers, even though the government's ICT knowledge is insufficient. When suppliers do warn the government about potential problems, their warnings were often not taken seriously. Parliament, meanwhile, fails to fulfill its oversight role due to a lack of interest and expertise in ICT. This lack of effective oversight leads to a significant waste of taxpayer funds. While it is not possible to determine the exact amount that has been wasted on ICT projects since 1995 (the last time a financial overview of ICT costs was conducted), experts estimate that the waste could be as much as 1 to 5 billion euros per year.

To address these problems, the committee recommended the creation of a temporary ICT supervising agency. This agency should be small and decisive, and its employees should be independent experts who can evaluate the chances of a project's success based on their knowledge, experience, and expertise. They should be able to provide advice on changes that could increase the success and effectiveness of a project, and they should have the authority to decline a project if it is deemed unnecessary or unlikely to succeed.

According to this committee, it would not be possible to completely solve all of the problems with the current system. Although a few well-designed and consistently implemented organizational measures could address a significant portion of the identified issues. However, if only some of the recommendations are implemented, the committee expected that the government will continue to struggle with ICT and waste taxpayer funds.

In the context of a different project, Ton Elias commented on this subject again in 2022 [50]. According to Elias, the ICT authority established to ensure that ICT projects met basic requirements was often bypassed and not independent enough to “effectively monitor and block misguided plans from the House of Representatives” [50]. Furthermore, Elias reported that the House of Representatives did not act upon the reports issued by this authority. As an example of this lack of progress, Elias cited the development of software for online video conferencing, which was given to a foundation created by a public official involved in the project without a public tender, at a cost of nearly 900,000 euros.

4.2 Bureau ICT-Toetsing / Adviescollege ICT

In response to recommendations from the parliamentary inquiry stated in Section 4.1, a temporary ICT supervising agency called the BIT was established. Based on the specifications stated in this inquiry, the organization was designed to be completely independent, operating as an extension of the Ministry of Internal Affairs. The organization itself consisted of a "Bureamanager" who approved reports, and a small team of permanently employed experts and temporary experts from the government and public sector as needed [64].

The agency's primary function was to analyze the probability of success for ICT projects and provide recommendations for changes. Importantly, all reports produced by the BIT were made public, providing transparency in government projects. An ICT project was defined as any project with an ICT component valued at 5 million euros or more and must be for a Ministry, an independent administrative body [65], the Council for the Judiciary, or the national police. The increased transparency provided by the BIT's public reports led to improved decision-making, as members of parliament used the information in the reports to inform their decisions. The reports also provided a third level of oversight for issues that were not identified during the initial and second checks.

A major challenge in evaluating projects is the requirement that the BIT can only examine a project if the minister in charge reports the project to the organization, or if it is requested by parliament. Legally, ministers are obligated to do so, but there is no penalty for failing to report a project. In practice, this means that not all projects that meet the requirements are reported. The advice given by the BIT is non-binding but is nearly always followed in practice, as the failure of a project after ignoring advice that could have prevented it is a politically sensitive situation. According to an evaluation of the organization, 65% of the bureau's advice is followed to the letter, while 35% of the advice is partially followed [66].

4.2.1 Political involvement

The effectiveness and power of the BIT are a major concern for high-ranking government officials. [67]. The organization was initially established because civil servants and politicians are unable to critically evaluate their own projects, and the BIT has proven to be a highly critical organization, sometimes recommending the complete shutdown of major projects. A former high-ranking government official stated that "every project that has been stopped results in a high-ranking official with a blemish on their career, and the more projects that have been stopped, the greater the administrative opposition." [68].

At first, this opposition was limited to delayed payment of bills and minor disruptions to work, but in late 2018, Secretary-General Maarten Schurink of the Ministry of Internal Affairs attempted to interfere with an independent report on the effectiveness of the BIT, which he deemed too positive. As the organization was only established for a temporary period of 5 years, this report would play a significant role in the decision to make it permanent. It is also not uncommon for ministries to try to influence independent reports, as seen in the political involvement at the "Wetenschappelijk Onderzoek- en Documentatiecentrum" (Scientific Research and Documentation Center), which is supposed to conduct independent scientific research. [69].

Based on this conclusion and correspondence with other government officials, Secretary of State Raymond Knoops of the Ministry of Internal Affairs reported to parliament in March 2019 that he would not draw conclusions from the independent report, and instead conduct their own investigation [70]. As the initial term for the BIT was only 5 years, ending at the start of 2020, this investigation would likely not be completed before the organization would need to be disbanded, although the deadline was later extended to the end of 2020 [70].

Based on the significant political involvement and the continued need for the BIT, it was decided in 2021 to make the organization permanent and rename it the "Adviescollege ICT Toetsing". However, there was no or no significant increase in either their budget or their number of staff. This shows the importance of the organization, while the lack of additional funding may indicate that not all issues have been resolved.

4.2.2 Evaluation of the organization

Since the AcICT has been made permanent, it has replaced a significant fraction of its board members in a short amount of time. Which could lead to a loss of organizational knowledge and negatively impact the effectiveness of the organization. At the beginning of 2022, the AcICT consisted of a chairman and two out of four positions being filled. In June 2022, these remaining two positions were filled [71]. On the 1st of January 2023, the chairman and one of the previous board members were replaced by two new individuals [72]. As a result, none of the original members of the BIT will remain.

As of January 2023, these board members are supported by a total of 20 employees, divided over a secretary-director, a secretary, various ICT experts, and other support staff. The task of the board members is to validate the accuracy and completeness of any report that the AcICT publishes, as well as to give structure to investigations and point it in directions that have not been covered yet [73]. Given that all board members currently hold numerous additional positions, this might impact the number of projects that can be evaluated [74].

In 2022, out of the 138 projects that meet the requirement of having an ICT component valued at 5 million euros or more, the organization has been able to finish evaluating 11 of them, with 6 under investigation. Previous years' output can be seen in Table 5 [75], which suggests a significant lack of capacity in the organization. With most requests unable to be reviewed. The Ministry of VWS in general has a very low number of projects being evaluated by the AcICT, with an overview of all of the ministries being depicted in Table 6. Notably missing any projects in 2021, during the pandemic.

Requests	2015	2016	2017	2018	2019	2020	2021	Total
Advice Requests	8	17	19	49	26	39	43	201
Published advice	3	12	13	15	15	10	14	82
Withdrawn		2	1			1		4
Denied				6	5	33	27	71

Table 5 - Overview of the Status of Requests at the AcICT

Ministry of ...	2015	2016	2017	2018	2019	2020	2021	Total 2015- 2021
General Affairs	0	0	0	0	0	1	0	1
Interior and Kingdom Relations	1	2	2	0	4	2	3	14
Foreign Affairs	0	0	0	0	1	0	0	1
Defence	0	1	1	2	1	1	1	7
Economic Affairs and Climate Policy	1	1	1	1	1	1	0	6
Finance	0	1	1	1	2	1	3	9
Infrastructure and Water Management	1	0	3	4	2	1	3	14
Justice and Security	0	2	0	2	1	1	1	7
Agriculture, Nature, and Food Quality	0	0	1	0	2	0	0	3
Education, Culture, and Science	0	1	1	1	0	0	1	4
Social Affairs and Employment	0	4	2	3	0	2	2	13
Health, Welfare, and Sport (VWS)	0	0	1	1	1	0	0	3
Total	3	12	13	15	15	10	14	82

Table 6 – Overview of the Number of Requests per Ministry

Given that the projected cost of the 138 qualifying and currently active projects is 6.203.340.000 euros [76], and that many projects are unable to be evaluated, it may be necessary to significantly increase the size of the organization. While an interview with an employee of the AcICT [73] and the Secretary of state of Digitalization, Alexandra van Huffelen [77], both stated that a sample-based approach would allow all projects to benefit from the results of published reports, this only covers one of the two primary aims of the organization. The first aim is to improve the management of governmental projects with a large ICT component in general, but the second aim is focused on each specific project. As their task is to provide advice on changes that could increase the success and effectiveness of a project and have the authority to decline a project if it is deemed unnecessary or unlikely to succeed.

Given the fact that the AcICT is named as a highly effective organization and the fact that it is significantly cheaper to either change or stop projects at an early stage, it is likely that increasing the size of the organization to at least the point where they can evaluate all projects, they are legally mandated to, would result in a net saving instead of an increased cost. As many projects are either completely changed [78] [79] or stopped [80] [81] by the AcICT this effect already appears to be present. The latest budget published for the AcICT is the budget for 2021, which was 5.4 million euros [59, p. 16], which is just 0.1% of the annual cost of the projects they are mandated to evaluate. And significantly less than the value of a single canceled project. Although it should be noted that due to the Covid-19 pandemic, the AcICT was unable to spend its full budget due to understaffing and Covid-19 reducing the amount of budget that would have been spent on employee training and education [75, p. 16].

This case study is another example of a project the AcICT was not involved in. This is even though the value of the tender for the project, as well as the later significantly increased budget, would likely qualify for the 5 million in ICT costs requirement, the nature of the data being processed, and the fact that a previous attempt to create such a system, the replacement for Praeventis discussed in Section 4.5, was heavily criticized. A possible reason for this is that not consulting the AcICT would also prevent public scrutiny of the development process and security standards of the project. Or the fact that the Ministry of VWS tends to not report projects as they have among the lowest number of submissions in general. Or the fact that in other events related to corona, there is also a refusal to open themselves up to public scrutiny, even when ordered to by the court [82].

4.3 Autoriteit Persoonsgegevens

The Autoriteit Persoonsgegevens is the Dutch Data Protection Authority and is responsible for ensuring that personal data is handled securely and responsibly in accordance with the General Data Protection Regulation. As an independent administrative body, the AP enforces the GDPR in the Netherlands and provides guidance and advice to organizations on how to comply with these regulations. However, the AP does not have the mandate to independently investigate matters. Instead, their investigations are initiated through reports from organizations, employees, or if a Data Protection Impact Assessment report containing high risks is issued by an organization and this organization chooses to involve the AP. The AP can also act in the event of becoming aware of a personal data breach that has not been reported, as this is a violation of GDPR. The nature of data breaches and the use of DPIA reports will be discussed in more detail in subsequent sections.

While the AP has the responsibility for ensuring compliance with GDPR and is responsible for delivering sanctions, its power in ensuring compliance is limited. Taking action relies on the reports they receive and improvements rely on the degree to which their advice is followed. Even with close involvement, the AP has limited capacity and cannot constantly monitor all data controllers. The AP believes that companies and organizations are primarily responsible for complying with the rules for protecting personal data but acknowledges that high-risk processes may still occur [83]. In these cases, the AP may take enforcement action at its discretion.

The effectiveness of the AP is further limited by a lack of manpower and other resources, which is a problem that may also be present within the AcICT. Research by KPMG on behalf of the Dutch government indicates that in order to fulfill its mandate, the number of AP employees needs to be increased from 182 in 2021 to 470 full-time equivalents (FTE) by 2025 [37]. However, it is important to note that due to the European Union's digital decade initiative [84], the AP's duties will expand to include several new data protection laws, such as the Digital Services Act, the Digital Markets Act, the AI Act, the Data Act, the Data Governance Act, and the Interoperability Act. This would suggest that an even further increase may be required in addition to the currently suggested increase.

While two separate motions to increase the AP's budget following KPMG's recommendations were approved by a wide majority in the parliament in February [85] and April [86] of 2021, the national government's budget for 2022-2023 [52] only provides a slight increase from the previous year's budget [87]. These figures for the period of 2021-2025 can be seen in Table 7.

	2021	2022	2023	2024	2025
Initial budget (2021)	26.274	25.075	25.076	25.077	24.946
Renewed budget (2022)	26.257	29.020	34.478	37.829	40.694
KPMG plan	-	44.000	53.000	57.000	66.000

Table 7 - Annual budget of the Autoriteit Persoonsgegevens (measured in thousands of euros)

4.3.1 Data Protection Impact Assessment

A DPIA is an instrument used to describe the data processing, assess the lawfulness of the data processing, determine the risks involved, and then take measures to prevent or reduce the possible negative consequences to an acceptable level. It is legally required to create a DPIA report and submit it to the AP if there is a probable high risk when processing the personal data for a given project. However, this is not limited to “risks to the rights and freedoms of natural persons” [88, p. 5]. In practice, the emphasis is often more on the risks for the organization instead. Examples of these are reputation damage, loss of customer confidence, loss of turnover, loss of market value, fines, and compensation/litigation costs [88, p. 5]. This thesis applies to public organizations; therefore, a loss of turnover or loss of market value does not apply. The term customer also does not apply in the traditional sense, while a service is still being received from the organization, an individual does not generally pay directly for the goods or service received.

I interviewed a data protection officer (hence ‘DPO’) at one of the ministries to determine the effectiveness of this process and to determine the exact method by which the government creates DPIAs [89]. A DPO does not create the DPIA but is instead tasked with evaluating them. In his organization, DPIAs are created based on the NOREA DPIA framework, which describes exactly which steps need to be undertaken [90]. While the DPIAs that have been published about this case do not use this exact framework [91] [92], the content, as far as it was not blacked out, matches the content of this framework.

A DPIA, based on the NOREA DPIA framework, consists of four sections. The first part is a systematic description of the (intended) data processing and the processing purposes. Which is a high-level overview of what the system's intended purpose is, as well as the scope and if new technologies need to be used. It also includes an overview of all data flows as well as which systems are involved in these data flows. It also needs to be stated which data is used and for which purpose. This data is divided into various categories, with some categories such as special personal data as specified in article 9 indicating a high risk [90, pp. 4-9]. The process owner is responsible for this section.

The second section consists of an evaluation by the DPO. Which evaluates the stated reason for the processing of data from the previous part. A DPO also needs to determine, based on the principle of proportionality, whether the purposes of the processing are proportionate to the infringement of privacy of the person concerned. As well as if the rights of the data subjects are protected, which includes informing individuals about the data processing method, the right to give or deny approval for processing, the right to request information from the organization and the right to have this information deleted [90, pp. 10-12]. The DPO is responsible for this section.

The third section consists of an evaluation of the risks that the system introduces. However, there is no specific technique that must be used to do this. Although such an evaluation generally consists of three parts which are risk identification, risk analysis, and risk evaluation. Throughout this process, risk treatment is applied to plan and implement risk treatment methods to minimize the risk where possible. If after this process there remains a ‘high risk’ and the organization either does not or cannot take additional measures, the AP must be consulted before processing starts [90, pp. 12-19]. It is not stated who is responsible for this process.

The fourth section is a legal document involving the various signatures of the actors involved in this process and the project in general. These are the previously mentioned process owner and the DPO, as well as a Chief Information Security Officer and the General Director/Board of Directors. The process owner signs that the information as described in the first part is correct and complete and that “organizational causes have been identified and the described organizational measures sufficiently mitigate the risks”. The DPO sings that the reasoning from part two has been found acceptable and that the most important risks have been identified. As well as that the DPO has been consulted and that his recommendations have been included in the DPIA. The CISO signs that “the described information security measures sufficiently mitigate the risks”, that he was consulted for his advice, and that his recommendations have been included in the DPIA. The director or board of directors signs that they are aware of the described remaining risks, that they have been accepted and that budget will be freed to implement the specified measures. [90, p. 20].

The DPO I interviewed however contrasts with this information. According to him, while a data protection officer is responsible for overseeing an organization's data protection strategy and implementation, he does not have the power to stop projects. While advice may be given, there is no requirement that this advice will be applied to the project. Neither is the DPO able to report projects to the AP. These decisions fall to the director or board of directors. Therefore, this DPO believes that DPOs and DPIA should not be expected to ensure that personal data is protected. However, as the DPO that was interviewed is employed at a different ministry than the Ministry of VWS, this is not a confirmation that the Ministry of VWS neglected the advice of their DPO.

In the context of the Covid pandemic, the DPO repeatedly emphasized the fact that the decision to create a DPIA may not rely on its legal requirement alone. Based on the fact that no DPIA has been made for both HPZone and HPZone lite, he stated that this could be a political consideration in the name of public health. As such a document is a paper trail of an evaluation into the security of a system and if it were ever to become public that a system is not secure, this could have a significant effect on the willingness of individuals to interact with this system. Therefore, just because it is legally required to create a DPIA, does not necessarily result in the creation of one. However, as the DPO that was interviewed is employed at a different ministry than the Ministry of VWS, this is not a confirmation that this is the reason that no DPIA(s) was made.

4.3.2 Personal Data Breaches

The purpose of this thesis is to investigate the potential non-adherence of the corona systems used by the GGD to the General Data Protection Regulation and use this information to evaluate if a FAIR-based architecture could be applied to improve compliance. In order to properly understand the implications of this research question, it is necessary to first define what constitutes a data breach. According to the European Commission, a data breach is defined as "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data transmitted, stored or otherwise processed” [31, p. 34]. The nature and severity of a data breach may vary, depending on the type and value of the personal data involved.

The number of reported breaches has remained relatively stable since 2018, after a major increase in this same year. These numbers have been visualized in Figure 2 [93]. There is however a significant change in the cause of these breaches. In particular, the number of breaches attributed to hacking, malware, or phishing has increased by 25% in 2019, 30% in 2020, and 88% in 2021, accounting for 9% of all reported breaches. Additionally, there has been an increase in the targeting of organizations that process large amounts of personal data, such as the GGD [94]. This trend highlights the need for organizations to prioritize the security of personal data to protect against potential breaches. However, the focus of this thesis is not on improving external security.

The Netherlands consistently ranks among the top three European countries in terms of the number of data breaches reported. This is evident both in terms of the number of data breaches per 100,000 people and in absolute numbers. In 2020, the AP received more data breach reports per 100.000 people than the bottom 20 countries combined. And the AP received more complaints per 100.000 people than the bottom 10 countries combined, with only the Irish supervisory organization receiving more complaints [95, p. 4]. However, according to the data protection officer, this does not reflect poorly on the country's security standards. Instead, it is a result of the fact that the Netherlands properly reports most data breaches, regardless of their severity. For example, the majority of data breaches reported in the Netherlands are related to personal data being sent to the wrong recipient, either via mail or email. In 2020, this accounted for 66% of all reported data breaches [96, p. 3], while in 2019 and 2018, this figure was 67% [97, p. 5] and 63% [98, p. 4], respectively.

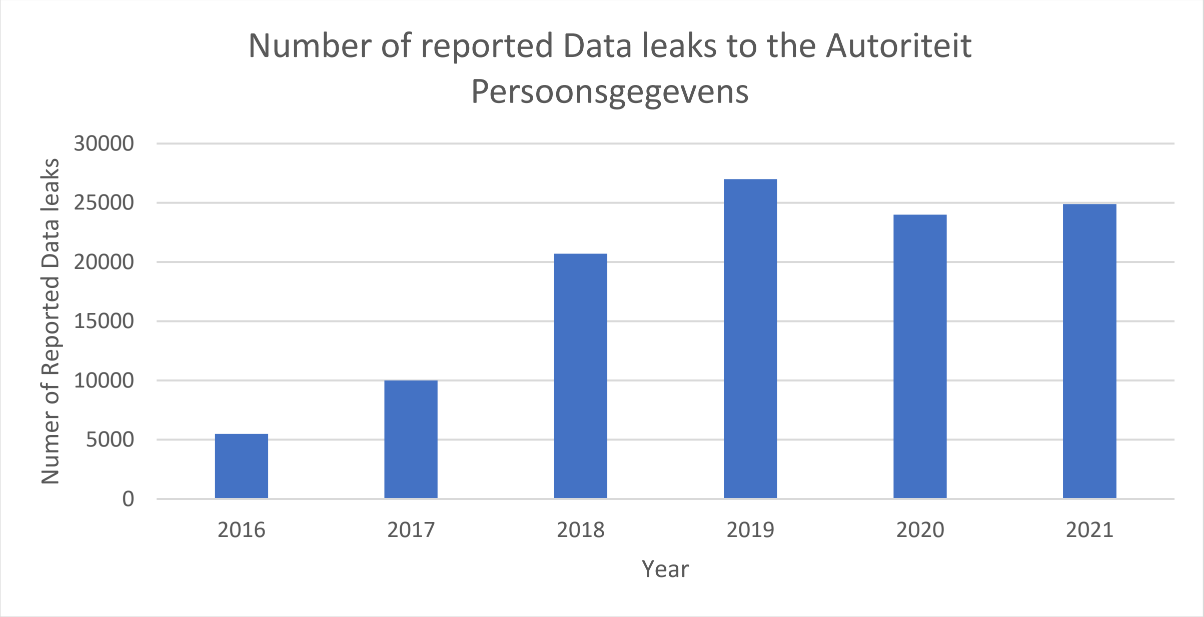


Figure 2 - Number of Reported Data Leaks to the Autoriteit Persoonsgegevens

It should be noted that the reporting method used by the AP, the only source of information on this topic, varies from year to year. Therefore, it is impossible to obtain comparable data for each year in each category. Of particular interest is reporting about the number of individuals affected by each data breach, which was only done in 2018. The majority of breaches affected only one individual (58%), or a small number of individuals (2-10, 21%). The data breaches that this thesis focuses on are significantly larger, affecting between 501-5000 individuals (4%), 5,001-100,000 individuals (2%), and more than 100,000 individuals (<1%) in 2018 [98, p. 6].

Two sectors are generally responsible for most personal data breaches: the healthcare sector and the financial sector [93]. The proportion of data breaches in each sector is shown in Table 8. While the number of personal data breaches in these sectors remained relatively constant, they significantly increased in 2021, which may be related to the Covid-19 pandemic. This could be due to an increase in the number of letters sent to individuals, which are responsible for the majority of breaches. It is also possible that some of the increase is due to the systems used by the GGD, but the available reporting does not provide sufficient detail to confirm this. The financial sector, on the other hand, has experienced significant changes, with the Public Administration sector overtaking it in terms of the number of data breaches reported in 2021 [99, p. 9]. The Public Administration sector is typically the third largest contributor to the number of reported data breaches.

Year	Healthcare Sector	Financial Sector
2017	30%	19%
2018	29%	26%
2019	28%	30%
2020	30%	22%
2021	37%	11%

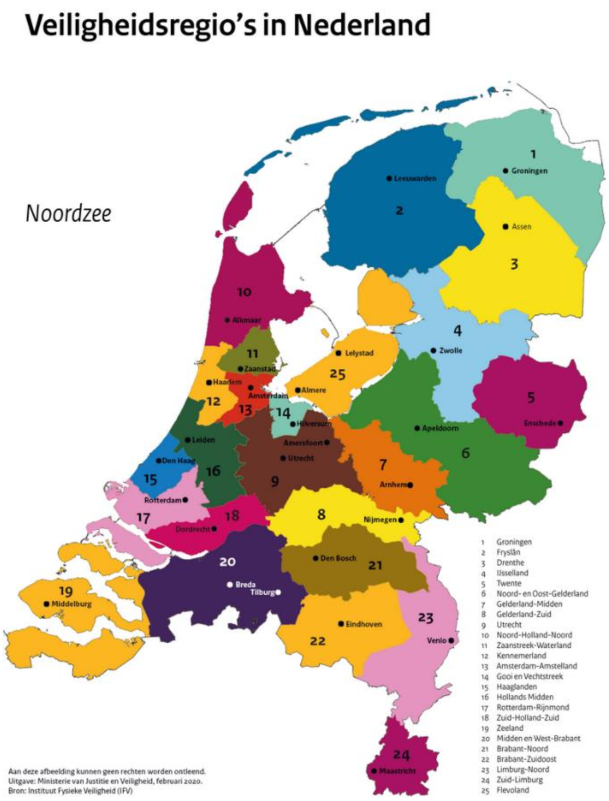
Table 8 - Number of Personal Data Breaches in the two most common occurring sectors

This thesis will not focus on addressing all data breaches, as most of them are not caused by data processing errors. A large proportion of data breaches are due to incorrect delivery information or a recipient moving to a different location. Additionally, the impact of most personal data breaches is relatively minor. Instead, this thesis will focus on addressing the larger data breaches, which can either affect a significant number of individuals or, depending on the measurement method, data breaches that affect only one individual but were caused multiple times by the same employee. For example, an employee may maliciously record a patient's medical records and repeat this process for numerous different patients.

Based on my interview with a DPO, data breaches are reported to the AP by the organization itself, instead of by DPOs, in the vast majority of cases. Due to the previously stated fact that most data breaches are relatively minor. These data breaches are not reported to the DPO, as they happen daily. Only in the event of significant data breaches will the DPO of the organization be notified.

4.4 GGD organization

The Gemeentelijke Gezondheidsdienst (hence ‘GGD’) is the public health care organization responsible for protecting, monitoring, and promoting the health of the inhabitants of the Netherlands. Under the “Wet publieke gezondheid”, their tasks include environmental medical science, health education, health monitoring, infectious disease control, and youth health care [100]. This case study will focus on the health monitoring and infectious disease control tasks of the organization.



Aan deze afbeelding kunnen geen rechten worden ontleend.
Uitgave: Ministerie van Justitie en Veiligheid, februari 2020.
Bron: Instituut Fysieke Veiligheid (IFV)

Figure 3 - Security regions in the Netherlands

The GGD is a federated organization that does not fall under any ministry or national-level organization, consisting of 25 GGD organizations, each belonging to one of the 25 “security regions” depicted in Figure 2. Each security region consists of a GGD organization, various municipal fire departments, the emergency services of that region, and the organization responsible for disaster relief. These security regions are controlled by a board consisting of the mayor of every municipality in the region, with the mayor of one municipality, typically the largest, serving as the chairman. The 25 chairmen together make up the Veiligheidsberaad, which discusses national-level affairs [85]. One of these chairmen is responsible for communicating with the national-level government.

The various GGD organizations communicate and coordinate with each other through GGD GHOR Nederland, which is the umbrella organization for the entire GGD. This umbrella organization has no authority over the various regions, instead, its responsibilities are limited to facilitating communication between the regions and executing projects developed for use in multiple GGD regions. Each GGD is managed by a ‘Directeur Publieke Gezondheid’, which together make up the “Raad van Directeuren Publieke Gezondheid”, with the director of GGD GHOR Nederland being the chairman of this council [101, pp. 6-7]. However, once again this is not a form of executive authority, but instead, a central point of contact for the national government and the Ministry of VWS to discuss matters. Actions themselves are taken by the GGDs, independently, in each security region.

Based on an interview with GGD personnel [102], which will be discussed in more detail in Section 5.4.5, the autonomy of the GGD proved to be a major problem for the national government. The Minister of VWS at that time wanted to take full control in combating covid-19 but did not have the authority to do so. As a result, different regions competed for funding from the state, for staff, and in the earliest stages of the pandemic even used many different COVID-19 registration systems. The interviewees even stated that the handling of this crisis threatened to be the end of the GGD.

The unique structure of the GGD and its place in the Dutch government make it more difficult to investigate this case as the “Wet openbaarheid van bestuur”, an instrument that can be used to request information from the government about their actions, only applies to the 25 GGD organizations in each region, and not to the umbrella organization of GGD GHOR Nederland [103]. However, In the context of the currently active ICAM lawsuit, GGD GHOR Nederland would coordinate the processing of WOO requests in each of the GGD organizations [86]. Even though releasing documents would be beyond the scope of this legislation.

4.5 Praeventis

As the COVID-19 pandemic is not the first instance of a disease requiring widespread vaccination, it is important to investigate the systems previously used to combat other diseases. Specifically, the systems used to support either the vaccination process or the testing and screening process to monitor the prevalence of a disease.

In the Netherlands, the most widely used ICT system to support testing and screening, and vaccination programs is called Praeventis, which was built in 2003 to support the Nationale Hielprikscreening, the Dutch National Immunization Program and the "Prenatale Screening Infectieziekten en Erytrocytenimmunisatie", and is maintained by the Rijksinstituut voor Volksgezondheid en Milieu (RIVM) on behalf of the Ministry of VWS [104]. The scope of this system is significant, as the immunization program alone is used to inoculate against 12 infectious diseases over 8 different appointments, for millions of people [105].

In 2014, Programma Vernieuwd Praeventis (hence 'PVP') was started to develop the replacement for Praeventis. Based on several external system investigations by different organizations and correspondence between VWS and the RIVM, it was determined that a new ICT system was necessary as the existing system was deemed to be not future-proof and too complex, consisting of 10,000 or more function points [106, p. 1]. Function points are a unit of measurement used to express the size of a system in software development. Subsequently, VWS (client) and the RIVM (contractor), in cooperation with the CIO offices of both organizations, began making initial preparations for a new system, including internal and external quality evaluations.

In 2018, this project was evaluated by the then-named Bureau ICT-Toetsing [106]. The BIT, based on its investigation, determined that there was no need for a new system as the existing system was stable and functioning well. Technological adaptations were possible, and there were no developments in the prevention programs that necessitated innovation. Adjustments that the RIVM deemed difficult to implement, such as ensuring GDPR compliance, had already been implemented [106, p. 3]. The BIT also determined that there was little chance of success for PVP. The program had not been able to clearly define the requirements for a new system, and employees only had an abstract image of the differences between the two systems. While flexibility and adjustability were named as requirements, there were no concrete plans to achieve these aspects, and legal requirements such as GDPR compliance and the "Wet zeggenschap lichaamsmateriaal" were not met. There was also no proper plan for the tender of this project, nor had any realistic scenarios been created.

The RIVM also lacked knowledge of software development and the required tooling, making it unable to properly evaluate tenders from companies offering their services in the creation of this system. This lack of ICT knowledge likely contributed to the decision to create a new system instead of improving the existing one. Instead, the RIVM should gain a deeper knowledge of the design and operation of the current software for prevention programs, allowing it to assess future improvements that can and should be made, properly review IT suppliers on a financial and technical level, and suggest alternatives. At the time, the RIVM had 1,660 FTE and an annual budget of 332.9 million euros [107].

The RIVM agreed with this assessment. Guaranteeing continuity of the systems supported by Praeventis was deemed the number one priority. Instead of a new system, improvements would be made to the current one [104].

4.6 Corona Melder App

While the focus of my case study is on CoronIT and HPZone (Lite), these projects were not the only projects that were developed by or for the Ministry of VWS during the Covid-19 pandemic. This section will discuss the Corona Melder app, an app designed to detect if a person has come into contact with an infected individual.

While these systems were developed in a similar amount of time, the development method that was used as well as the level of success were completely different. The development of CoronIT and HPZone (Lite) was conducted in a typical government project manner, where the project was acquired from a third-party vendor and then further developed to meet the requirements, following a tender in the case of CoronIT. The development of the Corona Melder app was done differently, being characterized by its complete openness regarding, among other things, design choices, source code [108], audits, and penetration tests of the software [109, p. 1]. Everyone was able to watch the development process and communicate with developers and stakeholders [110].

The project development of the Corona Melder App started using a hackathon structure where various entities would compete against each other in the creation of the best application [111]. On April 11, 2020, a tender was created on TenderNed for the development of smart digital applications to assist with source and contact tracing [112], with the Corona Melder app being related to 176 out of 756 applications for this tender. Of those 176 applications, 63 were selected for further review, and 8 of those were invited to participate in the public trial of the hackathon, with one withdrawing prematurely, resulting in 7 potential apps [109, pp. 7-8].

The hackathon itself was held on April 18-19, 2020, meaning that the entire process took less than 10 days. Based on the results of the hackathon, it was decided to create an internal app based on the hackathon results and further improvements. The entire development process was completed in a few months, and by August 2020, the app had been downloaded 500,000 times. An overview of this development can be seen in Figure 4 [109, p. 9].

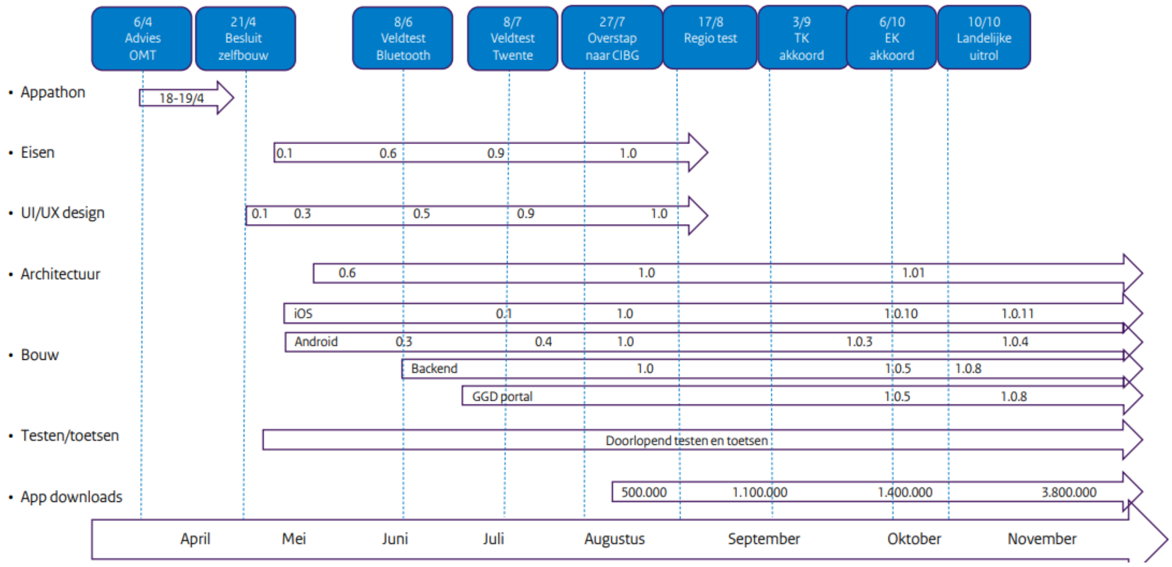


Figure 4 - Overview of the development of the Corona Melder App (numbers refer to versions of the apps)

Following questions raised in the first and second chambers of the Dutch parliament, the Secretary of State for the Ministry of Internal Affairs and Kingdom Relations requested that the AcICT evaluate the process of developing the Corona Melder app to identify lessons that could be learned for future projects. The aim was to also answer the question "What are the factors that have positively and negatively influenced the technical realization?" [109].

5. Problems that inhibited GDPR compliance at the GGD

The findings related to this research question have been organized into various subsections, with Section 5.1 providing an overview and more detailed description of the GGD's corona systems (CoronIT, HPZone, and HPZone Lite) and their use in the response to the Covid-19 pandemic.

The evaluation of these systems has been divided into three distinct periods, the initial design period, the actual development period, and the processing period. The first period, discussed in Section 5.2, focuses on the way the projects were set up and started, the actors involved in the development such as supervisory agencies and other governmental organizations, and alternative options to the systems that were considered but ultimately not used. The second period, discussed in Section 5.2, focuses on the technical aspects of development, such as system design and architecture. The third period, discussed in Section 5.4, focuses on the actual processing of data and the actors involved in this process, including internal information about the GGD during this period.

Section 5.5 assesses the response of both the GGD and the government to the problems encountered and described in the previous evaluation, with a particular focus on the personal data breach reported in January 2021. Which prompted the government to immediately take action after a period of more minor personal data breaches.

An overview of the potential GDPR violations arising from the issues identified in this chapter can be found in Section 5.6. Ultimately, only the enforcer of the GDPR in the Netherlands, the AP can confirm if the GDPR has been violated, in any of these areas. This thesis can also provide evidence for a probable breach.

5.1 Introduction

This thesis examines three systems used by the GGD during the Covid-19 pandemic, CoronIT, HPZone, and HPZone Lite. Although HPZone and HPZone Lite will be analyzed together for this thesis given that the differences between these two systems are minor. CoronIT was initially purchased and then further developed to manage appointments, register test results, and record individuals' vaccination status. This system was used throughout the pandemic, and remains in use as of January 2023, with the functionality to schedule vaccination appointments and register vaccinations added at the end of December 2020.

HPZone on the other hand was already in use by the GGDs for other diseases as their primary system for conducting source and contact tracing investigations. Initially, the full HPZone system, which included information from all infectious diseases, was used. Later on, HPZone Lite was developed from HPZone to limit access to only Covid-19 data and enable communication between different GGD organizations. This system was only used on a large scale in the early stages of the pandemic.

An overview of the testing and contact tracing process is depicted in Figure 5, with green blocks representing ICT systems maintained by the GGD, and the blue block representing systems maintained by the Rijksinstituut voor Volksgezondheid en Milieu (RIVM). Citizens can make appointments by contacting one of the GGDs, which are then recorded in CoronIT. The citizen then attends their appointment at a testing location, although as capacity increased, it became possible to go to a testing location without an appointment. The sample collected at the testing location is sent to a laboratory, which then sends the results back to the GGD. The results are stored in CoronIT and communicated to the citizen in question. If the result is positive, this information is passed on to HPZone Lite, which will contact the citizen to determine when they became infected and potentially identify the source of the infection. This step was only performed when it was still useful for tracing the source of infection, at the early stages of the Covid-19 pandemic. The aggregated data on positive test results and/or hotspots are then provided to the RIVM, which uses it to generate reports for the national government and GGDs.

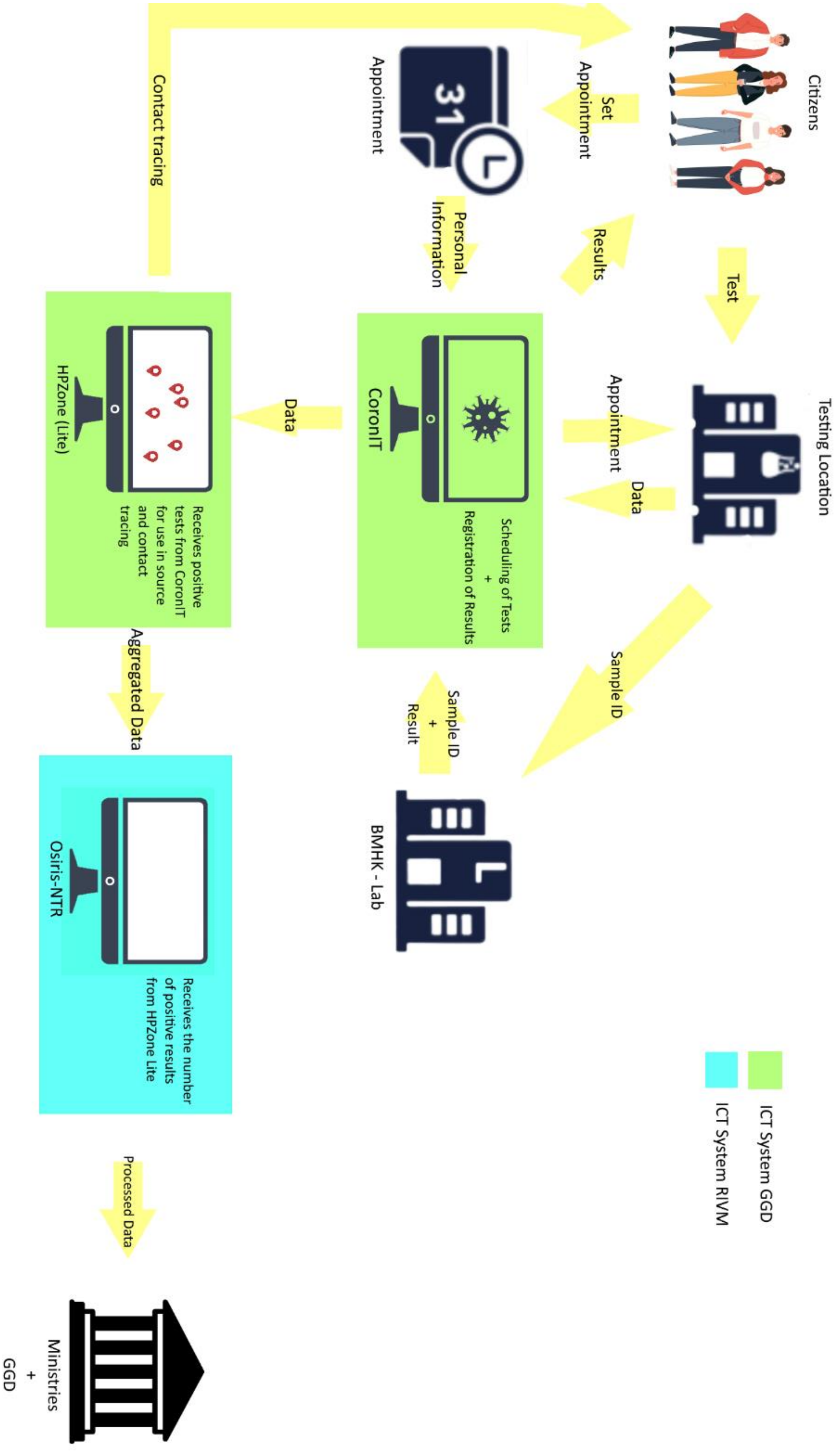


Figure 5 - Flowchart of the systems used for testing and contact tracing

In a later stage of the Covid-19 pandemic, it spread widely and contact tracing was no longer effective compared to vaccination efforts. This stage is depicted in Figure 6. In this process, citizens are invited to receive a vaccine dose based on their age and risk group, as determined by CoronIT. Once invited, they can schedule an appointment at a vaccination location. Later on, as capacity increased, it also became possible to go to a vaccination facility without an appointment. After receiving their vaccine doses, citizens receive proof of inoculation, which can be used in the Dutch corona app and the European Digital COVID Certificate. Information on the progress of the vaccination program is then shared with the RIVM and Dutch national government ministries.

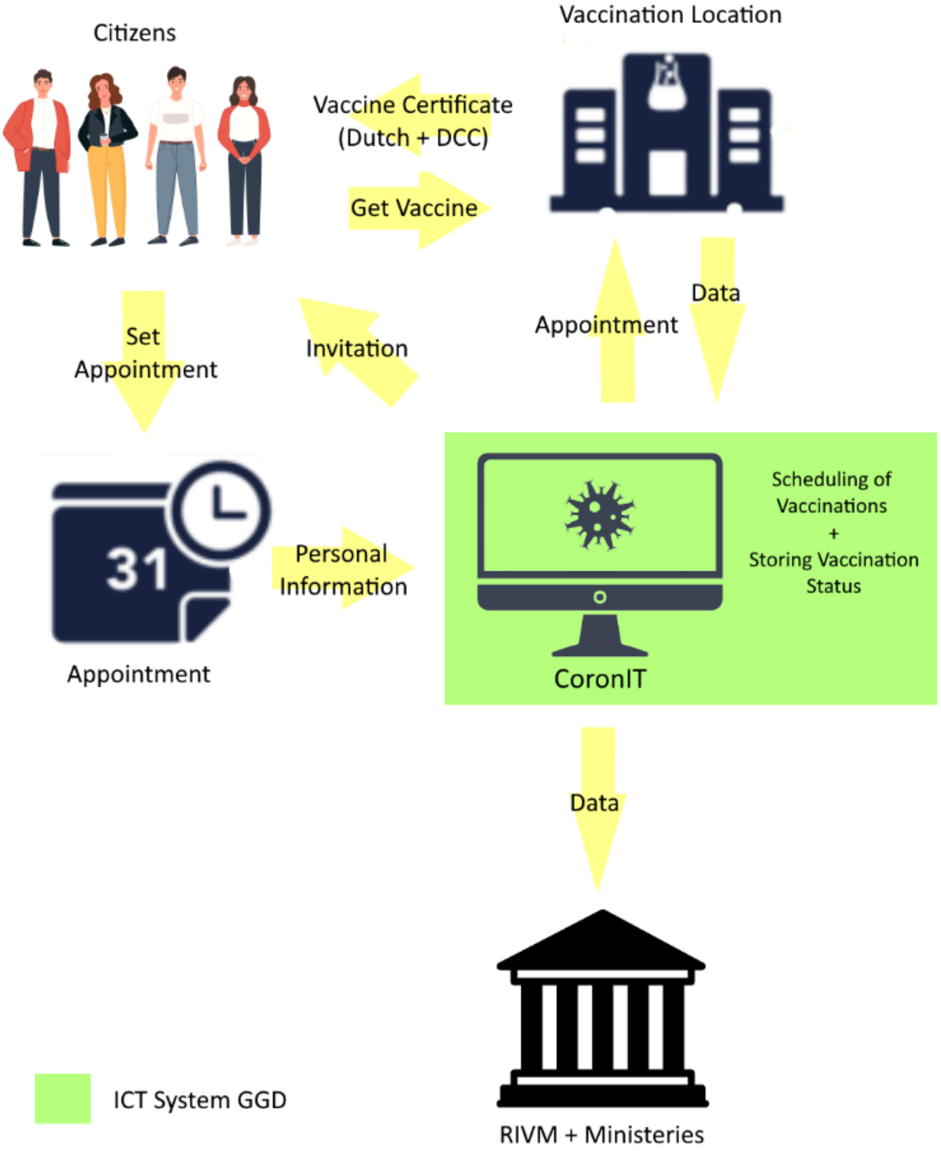


Figure 6 - Flowchart of the systems used for vaccinating citizens

5.2 Initial Stage

This section contains an analysis of the initial stage of CoronIT and HPZone and HPZone Lite. CoronIT may have been negatively impacted by the decision to have the project developed by GGD GHOR Nederland, an organization that may have lacked both the knowledge and capabilities to develop a national-scale system. HPZone and HPZone Lite were systems either already in use or based on a system already in use and were unlikely to meet modern security standards from the very start, which the GGD was aware of.

Both these systems had possible alternatives that were not selected in favor of developing CoronIT over making Praeventis useable in the Covid-19 pandemic and continuing to use HPZone and its derivative HPZone Lite over Go.Data, which was developed by the World Health Organization to use during a pandemic. Although the inherent insecurities of HPZone and HPZone Lite led to it being replaced by GGD Contact.

The role of the supervisory organization was non-existent, with the AcICT not being consulted as the projects were never reported to the organization and the AP not getting involved until reports of the major data breach in January 2021. The process around DPIAs had been carried out incorrectly, which resulted in no report being made to the AP. With no DPIAs having been made for both HPZone and HPZone Lite and the DPIA not being updated in the case of CoronIT. Although in CoronIT, the identified risk mitigation methods were either inadequate compared to the risk or carried out incorrectly.

5.2.1 CoronIT

In June 2020, GGD-GHOR Nederland was contracted to develop and implement CoronIT, at a cost of €15,814,700 excluding VAT and €19,135,787 including VAT [113]. The tender document for this project offers relatively little information, consisting of only four pages. It includes some functional requirements but does not address qualitative requirements such as data handling and storage, uptime, and so on. There is also no mention of vaccinations, as vaccines were still in development and it was unknown when they would be available for use on the general public.

The Ministry of VWS commissioned this project and no public tender was issued due to the urgent need for the system, which had to be ready for use within a very short period of time (eight weeks). Normally, the government would be legally required to offer national and European projects through the TenderNed announcement platform, but under article 2.32 paragraph 1 subsection 1 of the "aanbestedingswet 2012" [114], an exception can be made in a crisis.

GGD-GHOR Nederland was given the task of developing and implementing CoronIT due to the GGD's legal responsibility for maintaining public health and security, with GGD GHOR usually serving as the project developer for GGD projects. CoronIT was not developed from scratch as it was acquired from a commercial organization [115], although this thesis was not able to find the exact program or the vendor that sold it.

While this makes GGD GHOR Nederland a logical choice, several factors complicate this decision. As their role within the GGD is to facilitate communication between GGDs, they do not have a security region to manage and therefore have little experience with the practical requirements of such a system. Neither would they have experience with the technical requirements of such a large-scale system as developing and implementing an ICT platform on a national scale is a very different skill set compared to communication. This is worsened by the fact that the experience they have may prove to be detrimental as the projects used by the GGD before the pandemic would be designed for a limited number of specialized employees with full access to medical data.

In addition to a possible lack of experience to develop this system, a BIT assessment of the PVP (discussed in Section 4.5) determined that the RIVM, an organization significantly larger than GGD GHOR with experience in developing and maintaining national ICT systems in the healthcare sector, would not be able to successfully develop a new system. Therefore, even if GGD GHOR Nederland would have had the experience, the scope of this project would likely be beyond their organizational capabilities. The initial tender represented more than half of their annual budget of €30 million at the start of 2020, which increased to €296 million annually by the end of the year. This period of growth also saw the organization's workforce more than triple, from around 100 employees in March 2020 to 375 in December 2020 [35].

This level of growth may even damage an organization's capabilities more than it improves them as employees need training before they contribute to an organization. While there does not seem to be any Dutch-based data, US-based data suggest that an employee only becomes fully productive after 6 months, with 27% of employees taking more than 1 year to become fully productive [116]. Meanwhile, existing personnel needs to be used to train these employees, which reduces organizational capabilities. It should be noted however that this normally applies to adding a single or a small number of employees to a team, in the case of the GGD this would add almost 3 employees for every single employee. Making it likely that even more time is needed, and that reduction in capabilities is larger.

Furthermore, the unique structure of the GGD discussed in Section 4.4, introduces additional potential problems. In the context of this project, the status of the GGD as a semi-governmental organization rather than belonging to one of the ministries does not require them to make use of national government initiatives such as the four/five “overheidsdatacenters” [117]. Given that these data centers are set up to support large ministries with a significantly larger budget and storage needs, they are specialized in handling large amounts of highly sensitive data. It’s possible that through the use of these data centers, some of the issues with crashes, data loss, and general instability could have been avoided.

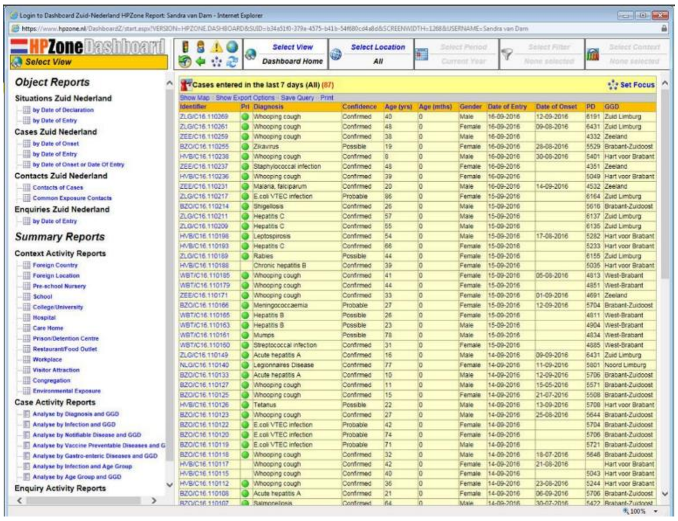
Finally, it is unclear why Praeventis would not have been further developed to support the processes that CoronIT was used for. The functionality related to testing in CoronIT is not unique, it’s a process used in many diseases and it matches the use of Praeventis as outlined in Section 4.5 where the main use of the system is to support national screening programs. According to the evaluation by the BIT, it already met all requirements of the GDPR and could be developed further. The only explanation this thesis could find is that this system would have been designed for use by professional healthcare workers, which would require a redesign for combating the Covid-19 pandemic. An explanation that in hindsight would not have made any difference as CoronIT faced the same issues that any such system would have. It could also have been easier to create a derivative of this system comparable to what was done for HPZone Lite, although that cannot be determined.

It was later decided to add functionality for vaccination planning and monitoring, after not being part of the initial system design. This is another potential flaw as including this in the initial system design would have given significantly more development time compared to the less than two months it got during the Covid-19 pandemic, where the planning of development started on the first of December 2020 [118, p. 15], with the first vaccinations on the 8th of January 2021 [119]. In the context of vaccinations, Praeventis was considered, however, the minister of VWS determined that would be easier and faster to modify the existing CoronIT system to support the vaccination program rather than relying on Praeventis. The minister further emphasized the risks of making such changes to a system that is in active use, as any interruption of service could disrupt the entire testing program, which potentially indicates a lack of ICT knowledge as any system should never be developed this way. Instead, it should be developed completely separate from the system that's in active use [118, p. 16]. The stated reason why Praeventis was not suitable is that it is not the system used by the GGD to schedule appointments, while CoronIT is based on the previous decision to not use Praeventis to schedule testing appointments.

This same document offers reasons why the choice to not investigate Praeventis may have been an incorrect one as based on the minister's statements that the government should not "try things we've never done with the vaccine for COVID-19, because it's more complicated than you think" and to stay with "executors with whom we also had experience and who also have experience themselves" [118, p. 16]. The only system used to support any vaccination program, certainly at this scale, would have been Praeventis, which has been used for every vaccination program since its creation. Given the BIT's assessment that the RIVM, a significantly larger organization with experience in managing such a system, would not be able to develop a new system to support vaccination programs in an unlimited timeframe, it is a significant assumption that the GGD would be able to do so in less than six weeks.

5.2.2 HPZone and HPZone Lite

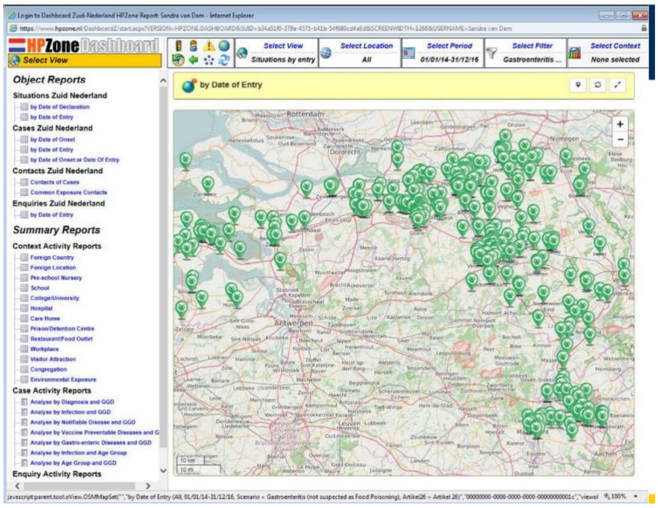
HPZone was already in use by various security regions to conduct contact tracing investigations and was also used during the Covid-19 pandemic to conduct contact tracing investigations into the Covid-19 virus. However, this system was designed to operate at a regional level rather than a national level, and for a small group of medical professionals. The creator of the system even stated that it offers no support during a nationwide public health crisis, defined as an infectious disease that exceeds the capacity of a single health organization [120]. This is precisely what occurred during the Covid-19 pandemic, as all 25 GGD health organizations were needed. Arnold Bosman, an epidemiologist who worked as an advisor for the GGD, describes the use of HPZone in this situation as "like using the Word writing program as a calculator and adding a chat function to it" [121]. Figures 6 and 7 show how the UI of HPZone looked as recently as 2015 [122].



Cases entered in the last 7 days (All) (87)

Identifier	Pri	Diagnosis	Confidence	Age (yrs)	Age (mths)	Gender	Date of Entry	Date of Onset	PD	GGD
ZLG/C16.110269	●	Whooping cough	Confirmed	40	0	Male	16-09-2016	12-09-2016	6191	Zuid Limburg
ZLG/C16.110261	●	Whooping cough	Confirmed	48	0	Female	16-09-2016	09-08-2016	6431	Zuid Limburg
ZEE/C16.110259	●	Whooping cough	Confirmed	38	0	Male	16-09-2016		4332	Zeeland
BZO/C16.110255	●	Zikavirus	Possible	19	0	Female	16-09-2016	28-08-2016	5529	Brabant-Zuidoost
HVB/C16.110238	●	Whooping cough	Confirmed	8	0	Male	16-09-2016	30-08-2016	5401	Hart voor Brabant
ZEE/C16.110237	●	Staphylococcal infection	Confirmed	48	0	Female	16-09-2016		4351	Zeeland
HVB/C16.110236	●	Whooping cough	Confirmed	39	0	Female	16-09-2016		5049	Hart voor Brabant
ZEE/C16.110231	●	Malaria, falciparum	Confirmed	20	0	Male	16-09-2016	14-09-2016	4532	Zeeland
ZLG/C16.110217	●	E.coli VTEC infection	Probable	86	0	Female	15-09-2016		6164	Zuid Limburg
BZO/C16.110214	●	Shigellosis	Confirmed	26	0	Male	15-09-2016		5616	Brabant-Zuidoost
ZLG/C16.110211	●	Hepatitis C	Confirmed	57	0	Male	15-09-2016		6137	Zuid Limburg
ZLG/C16.110209	●	Hepatitis C	Confirmed	55	0	Male	15-09-2016		6135	Zuid Limburg
HVB/C16.110198	●	Leptospirosis	Confirmed	54	0	Male	15-09-2016	17-08-2016	5282	Hart voor Brabant
HVB/C16.110193	●	Hepatitis C	Confirmed	66	0	Female	15-09-2016		5233	Hart voor Brabant
ZLG/C16.110189	●	Rabies	Possible	44	0	Female	15-09-2016		6155	Zuid Limburg
HVB/C16.110188		Chronic hepatitis B	Confirmed	39	0	Female	15-09-2016		5035	Hart voor Brabant
WBT/C16.110185	●	Whooping cough	Confirmed	41	0	Female	15-09-2016	05-08-2016	4813	West-Brabant
WBT/C16.110179	●	Whooping cough	Confirmed	44	0	Female	15-09-2016		4851	West-Brabant
ZEE/C16.110171	●	Whooping cough	Confirmed	33	0	Female	15-09-2016	01-09-2016	4691	Zeeland
BZO/C16.110166	●	Meningococcaemia	Probable	27	0	Female	15-09-2016	12-09-2016	5704	Brabant-Zuidoost
WBT/C16.110165	●	Hepatitis B	Possible	26	0	Female	15-09-2016		4811	West-Brabant
WBT/C16.110163	●	Hepatitis B	Possible	23	0	Male	15-09-2016		4904	West-Brabant
WBT/C16.110161	●	Mumps	Possible	78	0	Male	15-09-2016		4834	West-Brabant
WBT/C16.110160	●	Streptococcal infection	Confirmed	31	0	Female	15-09-2016		4885	West-Brabant
ZLG/C16.110149	●	Acute hepatitis A	Confirmed	16	0	Male	14-09-2016	09-09-2016	6431	Zuid Limburg
NLG/C16.110140	●	Legionnaires Disease	Confirmed	77	0	Female	14-09-2016	11-09-2016	5801	Noord Limburg
BZO/C16.110133	●	Acute hepatitis A	Confirmed	10	0	Male	14-09-2016	12-09-2016	5706	Brabant-Zuidoost
BZO/C16.110127	●	Whooping cough	Confirmed	11	0	Male	14-09-2016	15-05-2016	5571	Brabant-Zuidoost
BZO/C16.110125	●	Whooping cough	Confirmed	15	0	Female	14-09-2016	21-07-2016	5508	Brabant-Zuidoost
HVB/C16.110126	●	Tetanus	Possible	22	0	Male	14-09-2016	13-09-2016	5708	Hart voor Brabant
BZO/C16.110123	●	Whooping cough	Confirmed	27	0	Male	14-09-2016	25-08-2016	5644	Brabant-Zuidoost
BZO/C16.110122	●	E.coli VTEC infection	Probable	42	0	Female	14-09-2016		5704	Brabant-Zuidoost
BZO/C16.110120	●	E.coli VTEC infection	Probable	74	0	Female	14-09-2016		5706	Brabant-Zuidoost
BZO/C16.110119	●	E.coli VTEC infection	Probable	71	0	Male	14-09-2016		5721	Brabant-Zuidoost
BZO/C16.110118	●	Whooping cough	Confirmed	32	0	Male	14-09-2016	18-07-2016	5646	Brabant-Zuidoost
HVB/C16.110117		Whooping cough	Confirmed	42	0	Female	14-09-2016	21-08-2016		Hart voor Brabant
HVB/C16.110115		Whooping cough	Confirmed	40	0	Female	14-09-2016		5043	Hart voor Brabant
HVB/C16.110112	●	Whooping cough	Confirmed	36	0	Female	14-09-2016	23-08-2016	5244	Hart voor Brabant
BZO/C16.110108	●	Acute hepatitis A	Confirmed	21	0	Female	14-09-2016	06-09-2016	5706	Brabant-Zuidoost
BZO/C16.110107	●	Salmonellosis	Confirmed	64	0	Male	14-09-2016	30-07-2016	5422	Brabant-Zuidoost

Figure 8 - HPZone Dashboard Cases Overview



javascript:parent.tool.oView.OSMMapSet("","by Date of Entry (All, 01/01/14-31/12/16, Scenario = Gastroenteritis (not suspected as Food Poisoning), Artikel26 = Artikel 26)","00000000-0000-0000-0000-00000000001c","viewol

Figure 7 - HPZone Dashboard Geographical Overview

Given that HPZone contains information about all infectious diseases that require notification to the GGD according to Article 26 of the “Wet publieke gezondheid” [123], and that the user interface and user experience were difficult for inexperienced users, it was decided to create a version of HPZone, HPZone Lite, that limited data access to only Covid-19 data and improved the user interface and user experience to facilitate training. However, there is no difference on a system level and both systems can be seen as a single system using the same techniques, database, and source code [124, p. 26]. This new version was completed in August 2020. With this new system, large numbers of employees were able to more easily process and conduct source and contact tracing research. It also became possible for GGD regions to cooperate in cases of increased demand, which was not possible with HPZone [125].

A possible alternative to the use of HPZone was Go.Data, a free application developed by the World Health Organization. This application was specifically designed for use in source and contact tracing investigations during pandemics and was already in use in many countries for this purpose. However, the various GGDs decided to continue using HPZone because they were already familiar with it. Margreet de Graaf, the Director of GGD Frysland, stated that "You can only have a limited number of priorities. That is why using Go.Data was never seriously considered. The first wave was not the time to look around quietly and ask ourselves: maybe there is a better software package" [121]. The problem with this reasoning is that the people working with HPZone would not have been familiar with it, due to the significantly increased number of employees, most of which not medical professionals. Which eliminates the advantage of familiarity with existing systems.

In November 2020, a new system called GGD Contact was introduced to reduce the workload on GGD employees working in HPZone (Lite). The app allows individuals who have tested positive for Covid-19 to provide contact tracing information to the GGD by sharing lists of names and phone numbers of individuals they have interacted with [126]. This information is then entered into HPZone (Lite) by GGD employees. While the app aims to save time, the process of calling each individual on the list is still time-consuming. The app was initially planned for release at the end of December 2020 but was delayed due to practical testing at the end of January 2021 [127].

In February 2021, it was decided to replace the HPZone system entirely due to significant data breaches and problems with system performance. In part based on an investigation conducted from the 13th of January 2021 to the 4th of February 2021 called “Axis into ICT” [124, pp. 26-31]. This investigation concluded that HPZone (Lite) does not meet the legal requirements related to many standards and suggests that further use of these systems would not be allowed. If these systems continue to remain in use, there is a significant risk to their continued operation as a result of an increase in data with a negative effect on the image of all GGDs, GGD GHOR Nederland, and the public health service [124, p. 31]. At the time, two alternatives were identified: Go.Data, which had been considered as a possible alternative during the initial phase of the pandemic, and the creation of a new system called GGD Contact (system), which is distinct from the GGD Contact app [128, p. 2].

Prof. Dr. Jaap (J.T.) van Dissel, director of the Centre for Infectious Disease Control at the RIVM, expressed concerns about the decision to replace HPZone Lite on such short notice. He warned about the potential impact on the government's ability to "keep track of and understand the spread of the virus." Any new system will require extensive testing and development time, and should not only ensure security and privacy standards but also guarantee the same output. In the event of a hastily developed and insufficiently tested system with a completely new data flow, the RIVM would be unable to provide the necessary reporting and data to support the national Covid-19 strategy [129]. As the RIVM is dependent on data from HPZone Lite, the replacement process has significant implications for the data and advice the RIVM can offer in response to the Covid-19 pandemic [129,

p. 1]. The timing of replacing the registration system was highly unfortunate, as advice must be based on recent and complete data, and switching to a new system would require GGD employees to learn a new system.

For the transition to a new system to be successful, van Dissel stated that several requirements would have to be met [129, p. 4]. These include determining the release date of the new system in collaboration with the RIVM, ensuring that all data flows currently sent to the RIVM are functioning properly and that new data streams deliver the same variables and characteristics, providing the RIVM with at least two weeks to run both old and new systems concurrently, maintaining feature parity with the current system, and ensuring that data from the previous system is preserved.

While the decision has been made to choose the internal development of GGD Contact over Go.Data, this thesis has not been able to find the exact reasoning. Although GGD GHOR may have been able to gain enough knowledge during this process to successfully develop and implement such a system.

5.2.3 Role of the Supervisory Organizations

The role of the supervisory organizations, the AcICT (Section 4.2) and the AP (Section 4.3), has been limited in the context of the corona system used by the GGD. Neither organization was consulted during the design or development, nor were any external evaluation bureaus consulted to evaluate these systems. The AP only intervened when the various problems had been widely reported in the media in January 2021. After this breach, multiple other organizations were hired to investigate the failures of these systems, although this thesis will only be able to cover some of these, including the investigation discussed in the previous section.

Based on an interview with management-level GGD employees at one of the GGDs [102], problems were already known but it is unlikely that supervisory organizations, through “setting up a system of checks and balances”, would have been able to address them. This was simply not possible given the speed that was required to develop and scale up the systems and the lack of preparedness of the Dutch government for such a crisis situation. However, this also does not imply that no improvements could have been made. For such an improvement, they offered the example of notifying the central government themselves, that the regular process as present in the GGD would not be able to handle this crisis. Referring to “scaling up the number of GGD employees a bit”, which they believe is not a real solution.

They further stated that improvements are most urgent at the earlier stages of project management, which could be performed by the AcICT given that they had been consulted. Or it could have been identified if the process around DPIAs had been followed correctly. In their view, the government should realize that more “knowledge, budget and resources are needed to develop and maintain systems”. Which was lacking, but changed during the Covid-19 pandemic, where funding was significantly increased.

5.2.3.1 Adviescollege ICT Toetsing

While the AcICT is not responsible for ensuring the GDPR is enforced, GDPR compliance is an evaluation criterion that’s assessed before and during projects. In other governmental projects of this scale, the AcICT would typically be consulted to give their evaluation of the project itself and the process of its development. While this thesis is unable to determine the exact financial expenditure of the project, a significant portion of the total costs can likely be classified as ICT costs, given the nature of the project and the fact that at least an initial version of CoronIT was purchased from a vendor. The further development of CoronIT would only increase the ICT cost component. If the ICT component of the project exceeds 5 million euros, it would require the organization to report its project to the AcICT. However, no such report has been received by the AcICT.

According to correspondence with the AcICT, this is likely because GGD GHOR Nederland and the GGD are not national-level organizations nor any type of other organization which falls under the mandate of the AcICT [130]. However, the Ministry VWS, which is a national-level organization, falls under the mandate of the AcICT. And as this requirement applies to the client of any project as well, with the client of CoronIT being the ministry VWS, CoronIT falls under the mandate of the AcICT. As HPZone is a system that falls completely under the GGD and no tender for HPZone Lite could be found on TenderNed, these systems do not fall under the mandate of the AcICT. The fact that CoronIT has not been submitted could be due to the requirement that the minister of the ministry in question must report their project to the AcICT before it can take action, a requirement that is not always followed.

The AcICT did play a role in the Corona Melder App (Section 4.6), which was another project of the ministry VWS during the Covid-19 pandemic. Although in this case, the request to the AcICT was made by the secretary of state of the Ministry of Internal Affairs and Kingdom Relations and not the Ministry of VWS. After questions in the first and second chambers of parliament, the secretary of state of a different ministry, the Ministry of Internal Affairs and Kingdom Relations requested the AcICT to evaluate the process of developing the CoronaMelder app to be able to draw lessons from it for future projects. In addition to this process, the aim was to answer the question “What are the factors that influence positive and have had a negative influence on the technical realization?” [109].

It is unclear whether GGD Contact has been reported to the AcICT, however in a governmental document specifying its use at the GGD, there is no reference to such an examination [131]. It only refers to the examination performed in the context of the CoronaMelder App. As a tender for this system does not appear to exist, nor is any financial information related to the development of this system accessible to the public, this thesis can't determine if this would be required. It is also impossible to determine if the Ministry of VWS is named as the client. The DPIA from this system does list the Ministry of VWS as one of the definitions, but too much of the document has been blacked out to be able to determine this.

5.2.3.2 Autoriteit Persoonsgegevens

The AP is a reactive organization that would only become involved at this stage following an evaluation of a Data Protection Impact Assessment submitted by the GGD. However, there is no evidence that this has occurred with any of the systems discussed in this thesis. As noted in section 4.3.1, there is no legal requirement for the GGD to submit a DPIA unless the assessment indicates a high risk to data subjects.

According to statements made by the GGD GHOR and/or the ministry VWS in the DPIA for the GGD Contact system, a DPIA is required for nearly any system used by the GGD to combat the Covid-19 pandemic. While they initially state that any source and contact tracing portal is legally required to create a DPIA [112, p. 7], based on a list created by the AP [132], the reasoning they provide can be applied to a more general system as well. A system that processes medical and special personal data, and whose target and necessity imply "large-scale data processing," meets the requirement for a DPIA, especially in the case of an infectious disease such as Covid-19 where “large-scale processing of data (large amounts of data subjects (index and contacts) and their (contact) details) is a given” [112, p. 7].

A DPIA was created for CoronIT, however, there are various problems related to this document. While a DPIA is created at a certain point in time, it must be continuously updated to accurately reflect the current state of the system. If different personal data is processed or the extent of the collection and processing of data changes, either a new DPIA or an updated version of the existing one must be made. The latest DPIA to be released about CoronIT was published in June 2022 based on a WOO request from Stichting ICAM [91]. However, this DPIA makes no mention of vaccinations or the system's use in supporting the vaccination process. Indicating that this document was made somewhere between the start of development in early 2020, to November 2020 at the latest. Given the fact that the purpose or nature of data collection and processing has changed, as the process of vaccinating individuals is different from the process of testing individuals, a new or updated version of the DPIA should have been made. To confirm if this is the latest DPIA that was made and not an earlier version that would have been specified in their WOO request, I submitted my own WOO request, included in Section A1. This confirmed that this is indeed the latest version and no later version or different DPIA has been created.

A DPIA was not created at all for the legacy system HPZone, which was created before the implementation of the GDPR. While a DPIA would not be required for a legacy system, this only applies if there has not been any change to either the personal data that is being processed or to the extent of the collection and processing of data. However, such a change did occur during the Covid-19 pandemic, which applied source and contact tracing on a previously unprecedented scale, with employees that were not medical experts, targeting a significant extent of the Dutch population. A legal advisor, specialized in the GDPR, also agreed that this change would necessitate the creation of a DPIA [133]. To confirm if a DPIA has been made at some point in time, I submitted a WOO request, included in Section A2. Which confirmed that no DPIA has been created for this system during the Covid-19 pandemic.

A DPIA was created for HPZone Lite, however, there are some peculiarities. Based on the same WOO request as the previous paragraph, no such document has been made. Yet GGD Hart voor Brabant did publish a DPIA for HPZone Lite as part of the WOO request from Stichting ICAM [124, pp. 249-275], while this document is missing from the Woo request from Stichting ICAM published by GGD Zeeland. While all GGDs received this request, only these two have published anything on their website. The DPIA that has been published lacks the watermark that all other DPIAs created for GGD systems have, refers to it as an investigation done by GGD Hart voor Brabant [124, p. 250], includes a reference from February 2021 [124, p. 253], and refers to measures that the GGD took after the personal data breach in January 2021 [124, p. 272].

Based on correspondence with GGD Hart voor Brabant, this DPIA was indeed not made by GGD GHOR as part of the development stage of the system. Instead, this DPIA was part of an independent investigation started by GGD Hart voor Brabant and conducted by an external DPO after the personal data breach in January 2021. This is the first and only version of a concept product, dated the 30th of June 2021. However, given that GGD Connect would replace this system, it has not been discussed internally, nor was it shared with any of the other GGDs [134].

Given that HPZone Lite is a newly developed system, although derived from HPZone, it would not fall under the legacy system exception. As such, it requires the creation of a DPIA, even if HPZone would not. This is supported further by the previously stated fact that the replacement system for HPZone (Lite), GGD Contact both made one and specifically stated that source and contact tracing portal is legally required to create a DPIA and the fact that GGD Hart voor Brabant decided to make a DPIA for this system after the fact.

A DPIA has also been created for the Corona Melder App, which has also been continuously updated until the final version of the system [135]. The DPIA itself is publicly accessible in full, unlike the DPIA of CoronIT and the DPIA of GGD Contact. This possibly indicates that there is no requirement to hide information from created DPIAs. However, it should be noted that this is a different kind of system compared to the other two. In CoronIT and HPZone (Lite), intensive use is made of human resources while in the Corona Melder app, data is processed automatically. Eliminating many risks that apply to the processing of data with humans. The Corona Melder app also requires significantly less information, being designed to make user identification via the app virtually impossible. However, even in this instance, the DPO advises consulting the AP, "given the societal importance" [136, p. 2].

In an internal interview with management-level GGD employees in one of the GGDs, the interviewees suggested that some amount of political involvement took place around the process of DPIAs and their DPO, which will be discussed in more detail in Section 5.4.5 [102]. The interviewees believed it to be strange that the DPO in their organization would approve of the way and speed that these systems were scaled up, given that would normally "not even allow them to send flowers to the house of a sick employee as sharing their address would be a violation of GDPR". They further noted that the chairman of the data protection officers in their organization was a director in the department of the minister of VWS and suggested that the DPO was overruled in this instance. In addition to this, the DPO from a different ministry, discussed in Section 4.3.1, suggested that not making a DPIA could be a political decision, which would match what was discussed in the GGD interview. While this in no way confirms that political involvement played a part in this specific instance, this combined with the statement from the GGD that they were aware that HPZone would never meet modern security standards and the political involvement discussed later, also does not exclude it.

5.3 Development Stage

As discussed in Section 2.1, two core aspects of the GDPR are data minimalization and privacy by design and by default. Systems should be created in such a manner as to ensure that the least amount of data is being used to achieve the objective of the processing of data. And technical and organizational measures should be taken to ensure the data is processed safely at the risk to the data subject is minimalized. In this section, the technical aspects of CoronIT and HPZone (Lite) will be evaluated.

These aspects would normally be covered by a DPIA, however as discussed in Section 5.2.3.2, the DPIA for CoronIT was not updated and does not correctly reflect the state of the system at any stage beyond the summer of 2020 and the DPIA for HPZone Lite was created after the personal data breach in January 2021 had already occurred and included improvements that have been made based on that incident. Therefore, an evaluation of the technical aspects of the three systems is significantly more difficult. However, the lack of a DPIA for both HPZone and HPZone Lite and a lack of a current DPIA for CoronIT also already indicates a lack of awareness of both data security and risks. It's also important to note that HPZone and HPZone Lite have been replaced by GGD Contact due to it being deemed impossible to improve these aspects.

Combined with information from governmental investigations into this matter, various new reports, and an in-depth interview for a prolonged amount of time with two individuals who were actively involved at the management level of one of the GGDs in the Netherlands [102], this investigation was still able to get a clear picture of the technical aspect of these systems. This section will elaborate on what data was used in the corona systems and how data was processed, the security measures that have been taken, and the performance of the various systems as related to the availability and integrity of data. The level of detail depends on the availability of data, therefore while the information in this section is correct, it may not be complete.

5.3.1 Data Architecture

The three systems used by the GGD during the Covid-19 pandemic contain sensitive medical data on a large scale, containing highly detailed data over millions of different individuals. Given the nature and scope of this data, it is essential to ensure the security of this data, following the GDPR and various legal acts related to the protection of data, especially health data, in the Netherlands. Any violation of this may lead to significant fines from the AP and claims by affected individuals.

Citizens must be able to trust that their medical data is safely stored and only accessed, when necessary, given the privacy-sensitive nature of this information. Loss of trust in the security of medical data can not only impact the data subjects themselves but can also have broader public health implications. For example, if individuals do not trust the GGD to protect their data, they may be less likely to seek testing or vaccination, potentially endangering not just themselves but at-risk groups that rely on herd immunity as well.

The data in these systems was collected and used for a variety of purposes, including scheduling appointments, answering questions, providing test results, tracing infections, and recording vaccination status. All of these activities are crucial for maintaining public health. Based on the GDPR, it isn't necessarily a violation to gather such information as long as it is both required and the potential risks to a data subject weigh up against the benefits. However, this thesis will not attempt to determine if certain data should have been collected or not, the focus is on the way organizations manage this data.

An overview of the data collected and processed for use in CoronIT, for the testing process, is depicted in Table 9, based on its DPIA [91]. However, given that this DPIA has been made early in the development process and has not been updated, information from news articles and governmental documentation has been added. This included the date of birth. The “checklist of symptoms”, “Number of appointments at GGD locations”, and “Test result” fall under Article 9 of the GDPR, which specifies the conditions under which special categories of personal data may be processed. These data columns carry additional risks, which makes it even more important to take the appropriate technical and organizational safeguards. The BSN, as legally identifying personal data also creates a higher risk to the data subject.

The vaccination process was supported by CoronIT as well but as part of a separate database. This information has not been included in Table 9, but would likely include the number of vaccinations, the type of vaccination, the date of vaccinations, and information from the form that states if there are any expected medical implications such as passing out or an allergic reaction. These columns are also likely to fall under the special categories of personal data.

Personal data	Ordinary personal data	Special personal data	Legally identifying personal data
First name and surname	Yes		
Birth name partner (optional)	Yes		
Initials/nickname (optional)	Yes		
Date of birth (not included in the DPIA)	Yes		
Zip code	Yes		
House number	Yes		
Street name	Yes		
Place of residence	Yes		
Municipality	Yes		
Country	Yes		
Linked GGD	Yes		
Phone number (optional)	Yes		
E-mail	Yes		
Sex	Yes		
BSN			Yes
Barcode of sample	Yes		
Patient number	Yes		
Whether the person has worked in the last 2 weeks and if so, where	Yes		
Checklist of symptoms		Yes	
Number of appointments at GGD locations		Yes	
Test result		Yes	

Table 9 - Data used in the initial version of CoronIT based on the DPIA of CoronIT

An overview of the data collected and processed for use in HPZone (Lite) is depicted in Table 10, based on the DPIA that was published as part of an investigation into the personal data breach from January 2021 [124, pp. 249-275]. Unless ‘Infection Disease’ is a submenu that covers information such as which strain it is, the likely point of contact, and information about where the individual has been, it is unlikely that this table contains all information that was processed as part of the source and contact tracing investigations. There are at least four columns with special categories of personal data in this database, the “UZOVI – Number”, “Infection Disease”, “Ethnicity”, and “Country of birth”, although it is likely that the information missing from this table would also be classified under this same category.

Personal data	Ordinary personal data	Special personal data	Legally identifying personal data
BSN	Yes		Yes
BSN Validated	Yes		Yes
First name and surname	Yes		
Sex	Yes		
Date of birth	Yes		
UZOVI - Number		Yes	
Polis Number	Yes		
Address, place of residence and Municipality	Yes		
Infection Disease		Yes	
Occupation	Yes		
Ethnicity		Yes	
Country of birth		Yes	
Phone number	Should be here but left blank in the source		
Mobile Phone number	Yes		
E-mail	Yes		

Table 10 - Data used in HPZone (Lite), based on the later published DPIA of HPZone (Lite) (after January 2021)

This thesis has not been able to find any information about how this data was stored, in which format it was stored, how it was encrypted, or in what format it was anonymized when transferring data. However, while these aspects are important, the exact format is significantly less relevant than if these aspects have been carried out correctly.

For HPZone and HPZone Lite, these aspects were not carried out correctly, based on the previously mentioned “Axis into ICT” investigation [124, pp. 26-31]. There are no differences between HPZone and HPZone Lite on a system level, as they use the same techniques, the same database, and the same source code. The only changes are limiting access to just Covid-19 data and some kind of UI/UX improvement. The structure of the system was never designed for large-scale use, with no improvements being made to improve this aspect.

Due to this design and a lack of maintenance due to it being unclear who is even responsible for improving qualitative aspects of the system, many problems became apparent. There is a lack of validation and/or control functions for the database, which results in invalid data being accepted, resulting in inconsistent data. This has not been addressed except for periodically running scripts that correct certain software errors, which only address the symptoms. This has also resulted in the database approaching the limits of its storage, with the report not stating if something is being planned to address this.

To address this would require technical documentation, which does not exist. While features have been added throughout the years, at no point was this written down to make maintenance easier. Even worse, the system made extensive use of manual tests instead of automatic ones, which makes it more difficult to verify if feature parity has been achieved with the previous version and is likely to result in even more errors being generated when behavior has unexpectedly changed.

Such an investigation has not been done on CoronIT, although given that the system has not been replaced and HPZone has, it implies that no investigation concluded that CoronIT suffered from these same problems. While personal data breaches occurred, none of them were reported to involve anything related to encryption or data storage being compromised. Instead, actors that were authorized to access data in general, were accessing data beyond the scope of their task.

The flow of data in these systems is depicted in Figure 5 and Figure 6 of Section 5.1, with data being shared between CoronIT and HPZone (Lite) and shared with organizations beyond the GGD, such as the RIVM and the Ministry of VWS. From the example of the investigation discussed in Section 2.1.3, it can be concluded that data was also shared with the CBS. This thesis has not been able to determine in which form data was transmitted between the GGD systems and the GGD and other organizations, which could be either in the form of sending data files that are anonymized or by sending relevant results, such as the number of vaccinations, to an organization like the RIVM or the Ministry of VWS.

5.3.2 Data Access

This section will discuss both the way data access was given internally in the GGD and any organization allowed to process information for the GGD and the likely reason this was done based on internal information from one of the GGDs.

CoronIT and HPZone (Lite) made use of a different data approach at the start of their design, with HPZone (Lite) being designed to operate in a federated structure, with each GGD in each region being responsible for its own data. CoronIT was designed from the very start to operate as a single system, used by all GGDs and external organizations. However, it was later decided that HPZone (Lite) should be connected between regions, with each GGD being able to give access to employees employed at other GGD organizations, or external organizations.

Employees working in HPZone (Lite) had full access to all data at their GGD and any of the GGDs they received access to. This also included full access to all functionalities, Including, but not limited to, exporting all data from the database into an excel or pdf file, which was a holdover from HPZone [136]. Employees working in CoronIT also had full access, however, this access applied to data from the entire country. While employees can use CoronIT in different capacities, and for different roles, it was decided that full access should be given so that employees would be able to quickly handle additional demand in any region, for any role.

The GGD was internally notified of the security risks that this introduces and it even made the news in November 2020, but the GGD did not find it desirable to limit access to data: "GGD employees need to be able to view all files to be able to perform their work properly for all 25 GGD regions. Although we pay a lot of attention to this, it comes with a risk." [137]. Although they were aware of the risks, there is no indication that control measures were adequate to reduce these risks to an acceptable level.

For HPZone, this approach to data access can potentially be explained by the experience the GGD had with healthcare systems before the pandemic, which they then applied to a crisis situation where system requirements are completely different. The key differences of which are the significantly increased number of people that would now be working in these healthcare systems, as well as the fact that these people would most likely not be medical professionals. This problem is discussed in more detail in section 5.4.2.

5.3.3 Data Monitoring

The DPIA of CoronIT identifies two major risks related to the logging of user activity. The logging of user activity to access an overview of appointments has not been turned on and recommended that this is turned on, identifying this is a high level of risk that would be required to be submitted to the AP [91, p. 20]. Although given that this is the latest version of the DPIA, there is no information to confirm if this has been implemented. The second issue is that while information is logged, information is not actively periodically and automatically monitored, nor is there any policy determining how and by who such controls are carried out [91, pp. 20-21]. There is a high level of risk that information in CoronIT would be misused, but not detected [91, p. 21]. As none of the data breaches have been detected by the GGD, this problem had not been addressed in any later version of the system before the major personal data breach in January 2021.

The DPIA of HPZone Lite [124, pp. 249-275], although made after the personal data breach in January 2021, makes no mention of measures that have been implemented either to the logging of user activity or any kind of method in which this is reviewed. Instead, it stresses the importance of monitoring user activity so that unlawful processing is detected and handled. The monitoring situation is more unique with HPZone compared to CoronIT however, as each GGD manages its instance of HPZone in a federated approach instead of the combined system approach used by CoronIT. Ensuring that user activity is being monitored requires implementing measures in all of the GGDs, including any employees doing source and contact investigations as part of an external organization. However, no such agreements had been made and no uniform protocols had been created to ensure this, even after the previously mentioned data breach.

News articles provide various examples related to logging, however, they do not match the requirements set out in these DPIAs. The GGD stated that they were doing random and risk-based controls, although they would not “explain how we do this in the interest of the effectiveness of these checks” [137]. This in itself is already a dangerous statement as security by obscurity has been discouraged for some time now as “System security should not depend on the secrecy of the implementation or its components” [110, p. 15]. An example of one of their control measures was provided to the NOS by an employee, who stated that “Once every few months they have to share their screen through Microsoft Teams and then open the digital waste bin. The manager then checks to see if stolen data from the corona systems was in there." [138].

At the scale these systems were operating, with the nature of the data they processed, automated security checks are a critical requirement. However, this would not be possible to implement until March 2021 at the earliest. However, according to a report from the AP from November 2021, these automated control measures were still not implemented, with a daily manual check being used instead. Logging was done on a system-level basis, based on interactions with the system. However, in CoronIT, this data was only looked at in the event of a complaint or incident, and in HPZone and HPZone Lite, it is unclear if this was ever done at all [139, p. 6]. The limited time that these log records were kept was also a major problem, which will be discussed in the context of the most major data breach, in section 5.5.3.

This lack of supervision is not unique to the GGD however, as it is also present in other governmental organizations. One example of which is the Dutch Tax and Customs Administration, a major governmental organization under the Ministry of Finance with nearly 30.000 employees. This organization has been unable to determine which information has been leaked out of the system and by which employees. There are technical limitations, in part due to the organization consisting of 900 different, sometimes unconnected, systems that make it impossible to determine which address details, license plates, or financial information have ended up in the hands of criminals [140]. While the GGD only utilizes a limited number of systems, this example shares the same large number of employees that have access to valuable data, combined with a lack of monitoring systems to ensure that employees do not misuse their position.

5.3.4 Data Access Environment

To ensure that access to sensitive information remains private and secure, access must be limited to individuals that have the right to access that data. This section will discuss the measures the GGD took to ensure that outside actors would not have access to data from CoronIT and HPZone Lite. While there has been no report that a personal data breach occurred due to this aspect, it is possible that such an incident was either not detected or that such an attack was not attempted as it was already easy to access personal data in another way.

The GGD did not require company-issued devices to connect to the various corona systems, either internally or for outside organizations, although some GGD organizations would hand out laptops to some of their employees. However, the majority of employees, especially at external parties, were not using company-issued devices. There were also no stated requirements for devices that employees would use. Given that hardware is not standardized and managed by the GGD, various security vulnerabilities are introduced. Without organization-managed devices, it is unknown if the devices used meet all security requirements, nor would it be possible to force security updates. Software installed on the device may also impact the security of personal data as it could be infected by malware. Or the device could be more vulnerable to outside attacks [139, p. 4].

Connecting to the GGD corona systems was done using an URL instead of via a program that would create a secure working environment. This can introduce security vulnerabilities as measures taken in that secure environment could be bypassed in this manner. In combination with the lack of company-issued devices and clear requirements for any devices used to connect to the platform, this further increases the level of risk. The process of logging in, at least as recently as November 2021, does require 2FA, making it more secure [139, p. 4].

5.4 Processing Stage

In this section, the processing of personal data will be evaluated. The first subsection provides an overview of technical issues related to the systems. The other subsections elaborate on the main causes of problems that were identified in this investigation, which are the previous experience of the GGD, the level of growth during the Covid-19 pandemic, the agreements and cooperations with outside organizations, and the level of political involvement in this process. Together with the previously stated design of the system, this lead to numerous personal data breaches and other problems, until a significant personal data breach in January 2021.

5.4.1 Technical issues

Numerous technical issues occurred in the systems used by the GGD during the Covid-19 pandemic, as HPZone was never designed to be used in this capacity. CoronIT also suffered technical issues during the pandemic, although this was a newly acquired system specifically for this pandemic.

One month after CoronIT went live, in August 2020, CoronIT suffered a major system crash [141]. This occurred after 73,000 people had been tested and registered in this system in a week and resulted in the entire system becoming inaccessible for an entire day. As HPZone relies on data from CoronIT, this made it impossible to perform contact tracing in occupational groups as there was no information about test results, the number of tests done, or which percentage of people have been infected.

The cause of the crash was stated by Susan van den Hof, head of the Department of Infectious Diseases Epidemiology, as being related to the size of the database as it consisted of 370.000 data subjects. However, this explanation is problematic as while HPZone was not designed to handle this amount of individuals, CoronIT should have been designed with this in mind. As it was meant to be able to contain data about potentially every citizen in the Netherlands, which would be 18 million people, and numerous tests for each. In addition to this, a database with 370,000 rows of people, divided into various data columns should not be able to make any system crash as the total size would still be very small compared to databases in the business world. In those organizations, one would often work with millions to billions of data entries.

Crashes and interruptions remained a common occurrence after this incident, which is problematic as this data is used by the RIVM to create advice and policy for our corona strategy. For example, in October 2020, the number of corona infections according to the system had dropped below 9.000 a day twice in two days. However, this was entirely due to an ICT problem with the GGD, which meant that an incorrect count of the number of positive tests was sent to the RIVM. This number of positive results was then added to a later day, which meant that 4 out of 7 data in a single week had inaccurate counts [142]. This could potentially result in different actions being recommended.

In addition to crashes and interruptions, the level of activity on these systems would also cause them to slow down. Due to this, people would have to wait longer to make an appointment, which carries the risk of people giving up on calling and not getting tested anymore. In some situations, such as in June 2020, due to a massive influx of more than 323.000 callers, calls weren't even sent to a queue but would instead be disconnected to protect the phone line from not entirely collapsing [143]. The resulting number of untested positive individuals going out results in another danger to public health. Making it very important that even these technical issues are addressed.

5.4.2 Problems related to previous experience

Before the Covid-19 pandemic, the various GGD organizations were relatively small organizations with only a few hundred employees at most. These employees were divided into numerous departments of which infectious disease control, which CoronIT and HPZone would belong to, is just one. In the GGD of the interviewees, their department for infectious disease control consisted of 5 to 10 employees at the start of the pandemic. Systems in use by the GGD were designed with this in mind, as it could be expected that these systems would be used by small teams of medical professionals with a significant amount of expertise in this field. However, even this version of HPZone (Lite) did not meet many legal requirements, such as the NEN7510, NEN7512, ISO27001, and the GDPR ICAM [124, p. 29].

The extent of the pandemic required these systems to be deployed on a much bigger scale, with additional personnel that lacked the knowledge and skills of these specialists. These new employees found it difficult to work with these systems, not just because they were unfamiliar with them but also because these were complex systems with a difficult UI/UX. Which was further complicated due to significant time pressure. An example of this UI/UX, from 2015, can be seen in Section 5.2.2. While this example may be older, any improvement since then has not been enough as a 2022 investigation determined the UI/UX to not only be complicated in interaction but also did not meet the legal requirements for Web Content Accessibility Guidelines [124, p. 29].

In expanding and scaling up these systems, there was a general competence problem. The medical professionals, doctors, and nurses that used to be the only ones working in these systems wanted to keep control over them. They believed that other people were "interfering in their field" and that by working harder, they would be able to remain in control. However, medical professionals generally do not have software knowledge, so their knowledge did not translate to the ability to develop systems. The interviewees stated that this need to remain in control lasted for some time as it "took more than half a year before they really started to feel, yes, this is so big, we can't do that anymore". They also provided a possible explanation for this, as in the medical world, there is a great urge to keep things in their own hands, and not doing so would be a violation of medical confidentiality. A differently designed system does not necessarily violate medical confidentiality, but it would increase the level of risk.

The fact that systems were designed with small teams of medical professionals in mind and the fact that these same individuals wanted to remain in control of these systems can explain the design choices made in the Corona systems in use by the GGD. This applies especially to HPZone, as this system was already in use by the GGD and continues to be used. Medical professionals working in limited teams, with an NDA and a VOG, would not require systems to be designed with different levels of authorization in mind. Each user would require the highest level of authorization for his/her work and even the inclusion of an export function would be a required functionality.

However, in a proper system for the specific, and limited, duties of non-medical employees, access can and should be drastically reduced. The interviewees believe that there was no need for such detailed information in the context of what employees require to be able to do source and contact tracing investigations or to schedule appointments. The current system design stems from the "old primary thinking of the infectious disease doctor, who must know exactly what the patient is getting and which treatment because he must be able to account for it afterward".

In these newer systems, there was a significantly higher number of employees, and the average level of expertise was greatly reduced. In the case of HPZone (Lite), source and contact tracing is a medical act, which requires employees to follow training. This training, a theoretical e-module that can be completed at home on a laptop, takes roughly two days, according to the GGD. Followed by three days of practical training, under the supervision of nurses, doctors, or other experts which would naturally be very busy during this pandemic. For CoronIT, there are no such requirements.

5.4.3 Problems related to growth

In the early stage of the pandemic, it was uncertain if a vaccine would be ready to be distributed anytime soon, and the extent of the spread of Covid-19 was relatively minor. At this stage of the pandemic, the focus was on doing source and contact tracing to determine the exact source of the infection and the scope, to be able to limit any further spread. As each GGD would normally only have a few individuals that worked in this area, and these investigations could take as much as 8 hours per incident, it required the various GGDs to grow by a significant amount. In a later stage of the pandemic, large-scale testing required even more individuals to be hired.

According to the interviewees from the GGD interview, this process was heavily pressured by the Ministry of VWS to quickly increase the number of employees. Every GGD was required to provide weekly reports on the number of employees they were able to recruit at both a regional and national level. In their region, they experienced growth rates of up to 1.500% in the number of employees in a single week, with the only factor of importance being to supply an even greater number of employees. They further stated that “security at that time, as we are used to in the Netherlands, was really not a top priority”.

However, this unprecedented level of growth was not limited to their own region, occurring in every security region in the Netherlands. As a result, all GGDs were competing for the same population of employees. This was further complicated by the fact that while the GGD is charged with conducting source and contact tracing in their region, the national government was also hiring on a national level for the national organizations they created. The GGDs had no say over these national-level organizations and would even have to hand over employees they trained to these organizations, further increasing the number of employees they needed to hire.

In this entire process, the minister of VWS and the rest of the ministry were moving power away from the GGD, over to their ministry. In the opinion of the GGD interviewees, the minister of VWS “was not so much a director, as he was a project leader. Creating all kinds of national level organizations via his ministry, to appoint people there”. This created a high level of uncertainty as it became unclear what was being managed by the GGDs and what was being managed by the Ministry of VWS, nor was there any overview of what the capabilities of each organization were. The lack of a clear plan led to the GGDs having to do everything at the last minute, which was further complicated by the ever-changing strategy of the national government.

To illustrate this level of growth, all GGDs combined employed 12.000 FTE before the Covid-19 pandemic. And as previously stated, teams related to doing source and contact tracing were a fraction of this number with the number of employees employed in testing being nonexistent. During the pandemic, 59.000 additional individuals were hired to support the various systems discussed. In total, the number of FTE at the GGDs reached 51.000, with 32.000 employees or 20.000 FTE being involved with the vaccination process, 11.000 employees being involved in source and contact tracing, and 12.000 employees being involved in testing facilities or in call centers where they set appointments for the testing and vaccination process [144].

As previously stated, system security should be based on the number of employees and the expertise level of these employees. Any system, certainly at this scale, should apply a rigorous hiring process and limit access to employees. However, both aspects were not followed by the GGD.

Due to the level of growth that was required, the hiring process was simplified greatly. For example, any employee was required to submit a ‘verklaring omtrent gedrag’ (hence ‘VOG’), which is a governmental declaration that there are no objections to an individual acting in a certain role. However, this was not implemented securely, nor were the control measures adequate. This requirement had a grace period of six weeks, in which the employee could already start working before such a document had been submitted. Which introduced security vulnerabilities as people who would not be able to receive a VOG were not barred from using the system. Furthermore, an investigation by RL Nieuws identified several employees that were either never asked to submit their

VOG or only months later [145]. Another source to RTL Nieuws reported that hires were able to create their own account instead of having an account created for them, which he did numerous times without being required to submit a VOG [138]. Yet another source stated that he could immediately start working in the corona systems, for an external organization, after a two-minute phone call and without submitting a VOG [138].

While data was initially divided into each security region, based on the original design of HPZone that did not make a connection between various GGDs, this changed during the Covid-19 pandemic. While GGDs continued to be responsible for their region and each region had its own call center number, with its own testing and vaccination locations, data access became shared. And no attempt was made to divide data based on location, rank, or task.

Instead, it was deemed that the significant demand from the public necessitated giving each employee full access to all data, from all locations. This allowed the GGD to quickly fill in additional demand when there was a rush in a particular region or a particular task. Which was required at least in August 2020 due to a shortage of manpower, with GGD regions no longer able to trace all sources and contacts of corona patients. Major GGD regions such as Amsterdam and Rotterdam had stopped intensive tracing of the sources and contacts of corona patients, with GGD regions Groningen and Friesland aiding staffing shortfalls.

5.4.4 Problems related to external organizations

As stated in the previous section, external organizations were used to support the GGD during the Covid-19 pandemic. These organizations did not fall under the GGD, introducing various problems related to differing standards among these organizations. External organizations made use of employment agencies, which lead to further differences between employees hired internally by the GGD and people hired by one of the various external organizations and sub-organizations. The main differences consisted of the level of pay that employees would receive and the level and duration of training given to employees.

Internal employees at the GGD would receive 15 euros an hour, while external employees would only receive 10 euros an hour. This is a significant difference that influences both the quality of the personnel that would apply to the job, as well as an employee's motivation when they received the job. The compensation paid by external organizations is the same or close to the same as the Dutch minimum wage, which means that a job paying the same compensation can be found quickly. Something which can not be said for the salary paid by the GGD internally. A lower salary would also make it more likely for employees to cooperate with criminal actors to sell personal data. One external employee even stated that “I don't feel connected to the GGD at all, I only do this job because it's an easy side job and I need the money” [138], which is in clear contrast to the attitude of an internal employee that states that “I work here with heart and soul for our health” [138].

Internal employees at the GGD would receive training related to dealing with callers and working with the system, while external employees would be given training using a video connection, given by people who had little experience themselves. An ex-employee at one of these companies stated that “You work with a fairly complex system, with several programs next to each other that are all new. I didn't feel like I was ready, but I started anyway” [146]. Due to this lack of training, employees would seek help and share screenshots of the systems in group messages with other employees, often with clearly visible personal data.

In one of these hired companies, Teleperformance, a call center that provided 2100 employees for the scheduling of tests and contacts data subjects about negative test results, hundreds of employees had access to information they did not require for their job [147]. With a combination of a date of birth and a last name or a street name and a zip code or a BSN, these employees had full access to all records in CoronIT. Which included all test appointments, results, medical notes, and phone numbers of people they would have no reason to contact. This was reported in September 2020.

Employees at Teleperformance also mentioned that although they had to sign an NDA, they had received little to no training related to privacy. An ex-employee states that “Most people don't want to violate the GDPR, but it does happen”, as they simply lack the knowledge to know. If people that are currently waiting for their test results called the scheduling line, employees should state that they don’t have access to their results, even though they do. In some cases, callers did receive their test results, which goes against the conditions for data processing. Employees would also contact their family members and friends to inform them of their test results, outside of the proper communication channel.

5.4.5 Problems related to political involvement

While the structure of the GGD, being completely independent as discussed in Section 4.4, would suggest that there is no possibility of political involvement, this is not what happened during the Covid-19 pandemic. Instead, political involvement affected every single aspect of the way the GGD was utilized during the Covid-19 pandemic. From the hiring process to the amount of development time that was available, to the release dates of various functionalities, and even to internal decisions related to the design of the system.

Initially, the structure of the GGD and the central government's lack of authority in combating Covid-19 was a major problem for the minister of VWS, as he wanted to be able to implement measures directly. During one of the earliest clusters of Covid-19 in the region of the interviewees of the GGD interview, the minister of VWS repeatedly wanted to have the ability to directly steer policy and attempted to do so. However, it was pointed out to him that this is not legally possible as the authority to implement policy in this area does not lie with the central government or with the Ministry of VWS.

Cooperation does exist between the GGDs and the Ministry of VWS however. While all GGDs are independent, one of the 25 directors of public health is selected to act as the national representative. This representative meets with various stakeholders such as the Ministry of VWS, the RIVM, and other organizations to discuss the overall strategies and capabilities the GGDs can offer. Which is the context of this case would have been to combat the Covid-19 pandemic. The content of these meetings is then shared with all the other directors, as even this representative doesn’t have the authority to act or make any of the individual GGDs act in their region.

At the start of the Covid-19 pandemic, this approach led to certain problems. The various directors believed that this representative overstated the capabilities of the GGD and agreed with everything the minister of VWS, and by extension the central government, wanted. This also goes against the legal responsibilities of the GGDs in this area as the minister of VWS was attempting and able to steer policy. As a result of this, the situation became so unmanageable that it “threatened to be the end of the GGD”. Which was only prevented when this representative was replaced.

From this situation, it is clear that the current structure of the GGD and how we manage a health crisis do not match the intent of making the GGD autonomous, nor does it match the willingness of the central government to be able to implement action themselves. As a result, changing the current structure of the GGD in the area of infectious diseases is under consideration by the central government. This change would involve changing the responsibilities of the GGD, by moving the handling of acute threats and pandemics, such as the Covid-19 pandemic, to the national level while making no changes to the more local responsibilities of the GGD. There is already some legislation in the area of infectious diseases that allows for measures to be taken on a more centralized and higher level of government, however, this change would go significantly further than that.

According to the interviewees from the GGD interview, what was originally a health crisis had been turned into a political crisis. Moving far away from the "content and purpose" of combatting the pandemic, to a situation where the only matter of any relevance was "the wishes of the House of Representatives of the Netherlands and how decisions would reflect in the polls". And the center of our approach to the Covid-19 pandemic was the principle that "the minister of VWS was not, and is still not allowed to fail". To manage a public health crisis, you need to present a unified front able to quickly and decisively take action. Instead, any action was balanced according to "the position of the ministers" and the "wishes of the ruling and the opposition party". In their view, "if you have a common goal, which in this instance is fighting the pandemic, debates would not be required at all".

Due to the government being held responsible for the effectiveness of the GGD, as part of the way Covid-19 was dealt with, political involvement extended to internal design decisions related to the workings of the system. While GGD GHOR may have been the project developer, decisions were influenced or made by the Ministry of VWS and the central government. For example, the central government decided to not restrict data access to the GGD region of the data subject, instead, every data subject would be able to call every region and not be forwarded to the one they belong to. It was also decided to give employees the ability to look up all data related to every data subject, even when their responsibilities would not require this. While inherently insecure, this allows employees to fill in for demand in every region and every task, which would allow more people to be helped in a faster way.

The same influence can be seen in the hiring process, which would start following a call from the minister of VWS stating that they "needed to have a certain amount of people tomorrow, because the chamber of parliament reprimanded him, leading to another huge amount of people being hired". In this system, hiring decisions were not based on any kind of coordinated plan, with scheduling and measures to accomplish the hiring of a large group of individuals. Instead, it was reactionary, based on political issues, with regions competing against each other.

When asked if the interviewees believed that efficiency was put over security, they stated that they were not sure if this process was even efficient in the first place. Although it certainly was not secure as they believed there was no concern for the privacy of personal data, it had "nothing to do with data protection any more, this was the Wild West".

5.5 Organizational Response to Problems

While serious action was taken only after the major personal data breach in January 2021, the GGD was made aware of the problems related to the level of access offered to employees and the lack of security measures numerous times. While the resulting problems were relatively minor at the start of the Covid-19 pandemic, with personal data breaches being limited in scope, this changed as the pandemic progressed. With an ever-increasing number of employees and data entries, this only increased the dangers of lacking security practices. The number of internal warnings at the GGD increased until it resulted in this significantly larger personal data breach, with far-reaching consequences. This section will discuss the various personal data breaches, as well as how they were handled by the GGD.

In the summer of 2020, the GGD received dozens of reports of privacy issues surrounding the systems used to combat the Covid-19 virus. However, internal criticism of employees was waved away by managers, with problems being seen as not important enough, or not requiring a reaction. One GGD employee reported to RTL Nieuws that they “just don’t care” [138]. Although it remains a question why complaints would not have been made to the AP after their complaints went unheard at the GGD. In January 2021, the GGD stated that they were not aware of any such reports, instead they “appreciate employees who point out risks to us” [138]. Possible explanations for this are listed below:

- Reports were made to an external organization, which did not pass these reports on to the GGD
- Reports were made at one of the GGD organizations, which did not pass these reports on to GGD GHOR Nederland
- Reports were made to a manager, which did not pass these reports on to a higher level
- Reports were received by GGD GHOR Nederland, which chose not to act on them.

In November 2020, an anonymous whistleblower used the Dutch whistleblower platform, Publeaks, to report on the possibility and practice of misuse of the enormous database of the GGD. Stating that “Via the CoronIT Database, where the results of corona tests are registered, everyone can look up data from all kinds of people, acquaintances, and influential Dutch people”. This individual then proved this, by publishing information from two Dutch celebrities. Possibly in response to previous reports of potential misuse going unheard. This individual also reported that mobile phone numbers of handsome men and women who underwent a corona test were being shared [137].

Of these two individuals, one was Ahmed Aboutaleb, the mayor of Rotterdam, at that time under police protection due to various threats from criminals. The other person's identity was not reported. While Aboutaleb stated in response to this incident that there is no danger as he is a public figure and people already know where he lives, the same cannot be said for everyone else. An example of this occurred in a later incident when the personal data of Peter R de Vries and John van den Heuvel, two individuals who were being threatened and under police protection, was leaked [148].

It is important to note that the GGD itself was not aware of these breaches until it had been contacted by the AP, upon which they notified the AP of a personal data breach. These personal data breaches also did not lead to any action being taken to improve the demonstrated ineffective security measures employed by the GGD. Being seen as an acceptable risk.

In January 2021, two major personal data breaches were reported. The first data breach was reported by the NOS on the 21st of January 2021, at a private organization called U-diagnostics.

This organization performed a similar role to the GGD in testing individuals for Covid-19. The data breach resulted in the personal and medical information of at least tens of thousands of people becoming compromised and ending up in criminal hands. The clientele of the organization further increased the impact of this data breach as this organization was used not only by traveling organizations and football players but also by GP practices, healthcare institutions, and the Ministry of Defense for soldiers that have to be deployed [149].

The data released during this data breach is similar to the data contained in CoronIT, which was listed in section 5.3.1, in addition to travel destinations. The method by which data was compromised is also similar to the way this occurred in the GGD. Employees had full access to all information and made use of this to share phone numbers of famous dutch individuals, such as a model or a footballer for FC Utrecht. WhatsApp groups were also used to share personal and medical information without regard for privacy regulations. The difference in this incident is that passwords were being shared in groups, which were able to be used without any additional login requirements. Using these credentials, Nieuwsuur was able to get access to all information included in the database [150].

The second data breach was reported by RTL Nieuws on the 25th of January 2021, just four days after the previous report, at the GGD. This data breach affected both HPZone and CoronIT and consisted of the complete data of tens of thousands of Dutch people who had interacted with the GGD either in a source and contact tracing investigation or when getting tested for Covid-19. Individual data entries were offered for 30-50 euros each, with larger datasets selling for thousands of euros [151]. It should be noted however that like the previous incident, the personal data of everyone in the dataset was vulnerable and could have been affected.

As a result of this personal data breach, the Ministry of VWS was sued by Stichting Initiatieven Collectieve Acties Massaschade in February 2021, in the biggest class action lawsuit in Dutch history. They requested to receive 3 billion euros, for the 6.5 million people in the various corona systems. As the personal data breach could affect every individual in the database, they argue that every data subject is potentially affected and should therefore be considered a victim, no matter if their data has been used or not. Anyone whose data is stored in these systems should receive 500 euros as compensation by default and anyone whose data has been misused should be entitled to receive 1500 euros, which is significantly more than the 500 euros that the GGD only offers to people whose data has been misused [152]. Stichting ICAM suggests that by accepting this compensation individuals waive further claims for compensation, which the GGD uses to reduce the amount of compensation they have to hand out.

Experts however are unsure about this lawsuit, stating that the scope is too big. Professor of Mass Claims at Tilburg University Ianika Tzankova states that: “On the one hand, there is apparently an abuse here: the government must also adhere to rules. On the other hand, this is an example of a crisis situation. One could imagine that this is not the most sympathetic thing” [153]. As the lawsuit has yet to end as of January 2023, this thesis can not say if the legal system views this lawsuit in the same way or if the infringement of a data subject's rights to privacy outweighs the fact that this was a crisis situation.

5.5.1 Measures Taken

Based on the major personal data breach in January 2021, the GGD took (or planned to take) widespread action. Which included measures such as related to improving access security, authorizations, activity logging, restricting the search functionality, disabling the export- and print functionality, and notifying the victims of the personal data breach. In addition to carefully investigating HPZone and CoronIT in their entirety.

The GGD hired Fox-It, to do forensic investigations into the server-level loggings in CoronIT [154]. Based on the monitoring from Fox-IT, GGD GHOR hoped to be able to detect suspicious behavior in both previous logging information they had related to this data breach and to also be able to detect this in the future. This was done using a team that manually tries to identify suspicious behavior, 7 days a week. However, as the corona systems are used by tens of thousands of employees and the low number of employees Fox-IT has (350 according to their vacancy page [154]), with not everyone being involved in this process, it is unlikely that they would be able to monitor this effectively. Realistically, only automated systems would be sufficient, which was planned to release at the end of March 2021.

It is interesting to note that Fox-IT was also hired to perform extensive investigations into the security of CoronIT, specifically related to vulnerabilities in the system at the start of the project in 2020 [155]. Based on their investigation it was concluded that no additional actions were required. The registration and transfer of data had also been tested against the regulations regarding the protection of personal data, the GDPR. Although any such investigation would focus on outside protection, instead of internal protection. Therefore, it is unlikely that any problems related to system design concerning internal employees, specifically the decision to not restrict information and to not employ automatic monitoring systems, would be found.

For HPZone (Lite), the GGD disabled the biggest export functionality, and adjustments would be made for the other, required, export functionality. Access to these functions had also been limited to a significantly lower number of individuals, in a limited number of roles. Furthermore, HPZone would be transitioned out entirely, in favor of GGD Contact, as soon as the system would be completed. However, the GGD was already aware of the fact that HPZone would not be able to meet modern requirements, which is why it is surprising that no action was taken to either address this fact or start developing an alternative system while using this system as a last resort. The GGD implies this use as a last resort as they state that they used HPZone as the system used for supporting the testing process in March 2020 because it was the only system available, stating that "We made adjustments, but we also knew that a new system was needed.". This isn’t necessarily true, however, as they identified an alternative system specifically designed for use during a pandemic.

It's interesting to note that the GGD did not consider the data breaches to have any relation to “the failure, malfunction or possible insecurity of the system”. They blame the personal data breaches, not on systemic errors but on “malicious intent and the urge to make money off the backs of others”. They also attack the media, stating that they are spreading “stories full of facts and fabrications, inaccuracies and incompleteness, justified and unjust criticism” [155]. This could however be partially explained by a statement made by the GGD interviewees that stated that this is a problem related to the level of influence the GGD had, with internal systems being made by the Ministry of VWS instead of the GGD. Further stating that “If you aren’t the ones making the decision, it is ill-advised to appear like you are the owner of this problem and made the decisions yourself”. As such, the GGD would receive the blame for “the government’s decision to place speed and efficiency over quality and privacy by design” [102].

While these personal data breaches occurred due to internal employees, although made significantly easier through decision choices from the Ministry of VWS, the lack of any kind of automatic control system and a short data storage policy can only be attributed to the GGD. In addition to this, even if decisions influenced the GGD, the GGD is ultimately responsible for any processing done with their data, which includes being responsible for their internal employees and the employees of any hired organizations. As the organization was aware of the dangers from initial data breaches and of the fact that HPZone and HPZone Lite were inherently insecure from the start, changes, and improvements should have been taken earlier.

5.5.2 Evaluation of Measures Taken

After the major data breach, GGD GHOR had been under 'increased supervision' by the GGD [156]. Based on this increased supervision, the AP released an evaluation in November 2021 of the GGD, evaluating the implementation status and effectiveness of the previously mentioned measures and other improvements [157].

The AP concludes that while some of the announced improvements have been implemented, reducing the risk of personal data breaches, the risk of a personal data breach has not been eliminated. Significant risks to the protection of personal data remain, which will require additional measures. These measures are specifically related to addressing problems related to the significant number of employees and organizations involved with the processing of personal data related to testing, vaccinating, and source and contract tracing investigations.

Involved in this process are 25 different GGD organizations that each belong to a security region, GGD GHOR, six different national partner organizations (call centers and other external organizations), employment agencies, and IT suppliers. While this is already a complicated organizational structure, the AP investigation found that no clear agreements related to various security aspects from the various systems had been made. For example, related to access control related to the various regions and functions and the monitoring of logging information. As a result, it is unclear who is responsible for taking various security measures, which Increases the chance of new breaches in data security [139, p. 1].

The process regarding providing and removing access to employees was also implemented incorrectly, with there being no documented agreements relating to the assigning, editing, and revoking of authorization. Instead, authorization control for HPZone and HPZone Lite is provided by each GGD separately, with each GGD providing access to their system and assigning a role to employees and contractors working directly for the GGD. For externally hired companies, employees receive access based on their role, with a GGD organization sharing access to data from their region with this employee. If the employee starts working in a different capacity or for a different region, this access should then be revoked and replaced by a different one. There is also no documentation related to the rights and functionalities connected to roles. As a result, certain employees can access far more data than would be required based on their responsibilities [139, p. 5].

In response to various smaller data breaches in September 2020, GGD GHOR started the development of an automated control system to check log files, which would be finished in the fourth quarter of 2020. However, this system was still not in place at the time of the major personal data breach in January 2021. After missing this deadline, GGD GHOR would implement automated control systems in the form of a 'Security Information & Event Management' (hence 'SIEM') by March of 2021, a deadline that was still not met as recently as June 2021 [139, p. 6].

Since the data breach, the search functionality in CoronIT has been limited to no longer processing search queries consisting of only last names, a combination of last name and gender, or last name and date of birth. A similar change has not been made to HPZone however due to technical limitations at the suppliers. Instead opting to show less personal information in a search query [139, p. 7].

5.5.3 Size of the Data Leak

Determining the exact number of individuals affected by these data leaks will most likely be impossible, however, it is also unlikely that the official number used by the government is correct. With both HPZone (Lite) and the testing records of CoronIT being affected, any one of the millions of individuals in these datasets could be affected. Given that HPZone (Lite) contained a built-in export functionality, accessible to all employees, it's possible to rapidly download a significant number of data entries. CoronIT lacks this functionality as it is targeted to the individual, but examples further in this section will show that a significant amount of data entries could still be extracted from this system too.

RTL Nieuws informed GGD GHOR on the 22nd of January 2021 that a major personal data breach had occurred at the GGD, based on their investigation into datasets it acquired from criminal actors. These datasets consisted of the complete information of 600 individuals, which they verified by contacting a sample of these individuals with the contact information contained in these datasets. After verifying that these datasets were legitimate, RTL Nieuws was told by these criminal actors that they would be able to provide thousands or tens of thousands of additional records [158].

While RTL Nieuws informed GGD GHOR about the data breach, they did not share these datasets or their investigation with either GGD or the Dutch police, stating that they are an independent organization that does not function as an extension of any kind of investigation. According to the editors-in-chief of RTL Nieuws, the GGD itself is responsible for investigating personal data breaches and informing victims. They also deleted these datasets after publishing these articles [158]. This thesis was unable to determine if the national police or the GGD made any attempt to investigate this data breach using the same approach as RTL Nieuws.

Given that GGD GHOR stated that they were completely unaware of the personal data breach until it was reported to them by RTL Nieuws on the 22nd of January [124, p. 641], combined with the fact that logging data, at least at GGD Hart voor Brabant, was kept for 30 days [124, p. 646], it makes it impossible to investigate any case before the end of December 2020 from this side. As these flaws were present from the very start of its use in March 2020, this leaves a 9-month period where it isn't even possible to determine the number or scope of personal data breaches during this period. And it is unlikely that the few incidents reported in the news are all personal data breaches that occurred during this period.

Even for the period between the end of December 2020 and the 22nd of January 2021, it is unlikely that all personal data breaches would be able to be detected. The systems were never designed to be able to monitor user activity to such a degree, as the design only specifies manual checks, which makes the expected value of this logging limited. Given that these systems were used by tens of thousands of employees, with every interaction being logged, there is an enormous amount of data to go through. This can be significantly reduced through various intelligent queries but remains a significant amount of data to investigate.

According to a police investigation and the GGD itself, the scale of the personal data breach was limited to 1.250 data subjects. All of these individuals have been notified of the data breach and offered 500 euros as compensation [159]. To get this compensation, data has to be "unauthorized access, stolen and possibly sold", meaning that it's not a requirement for data to be sold, only illegitimately accessed. The GGD further states that "No evidence has been found for the alleged (large-scale) trade of data from the corona systems.", which contrasts RTL Nieuws' reporting that there are datasets of thousands to tens of thousands of individuals.

From an investigation of the limited number of criminals the police arrested related to these personal data breaches, it appears that all of them were overly suspicious to the point that it would be impossible not to identify them. In addition to this, an example presented further in this section would indicate that just a single individual would have been responsible for the majority of the data breach. Therefore, it is likely that the investigation done by either the GGD or the national police was only able to target the most blatant or least capable criminals in the small window of time they had while missing both criminals that took basic precautions to make themselves less suspicious and any criminal that accessed data maliciously any time before December of 2020.

This next example describes a personal data breach in CoronIT, which would be responsible for most of the total data breaches related to the GGD corona systems, and also shows that an export functionality is not a requirement for being able to leak large amounts of data. A 19-year-old man from the Hague was been accused in January 2022 of taking 795 pictures of dossiers from CoronIT and selling them for 1 euro per individual [160]. The police investigation into the matter was able to establish that this man logged into his account on a sick day, upon which he accessed hundreds of files multiple times. This is an example of a less capable criminal, not taking any basic precautions to avoid detection. However, this also shows the value of having any kind of automated monitoring system, as both the frequency of accessing files and the fact that a user was logged in at a time he should not have been working would be detected in even the most basic of systems.

Given the context of this particular case, it is unlikely that this individual was in any way linked to the criminal organization that offered a dataset to RTL Nieuws. The method was too obvious and inefficient and he also did not realize the value of the data he had. The suspect was approached by an individual that paid him for the data of people born in 1965 or earlier, as these are the people most likely to fall for phishing scams. The value of the data, 1 euro per individual, is in contrast to what RTL Nieuws reported earlier where this data was sold for 30 euros a piece.

A professional operation linked to a criminal organization poses a significantly bigger risk, especially if done in a way that circumvents any control measures. The easiest way to circumvent such a system is to construct attacks that are indistinguishable from normal behavior. Using such a method, even a sophisticated automatic monitoring system would not be able to detect an attack. An efficient way to do this would be to use a device that records your screen and uses OCR software to extract all personal information, which would expose the user to a significant amount of personal data while performing their job exactly as expected. If you convert the information into a format that does not resemble the GGD database, you could even sell the information as being from a completely different data leak.

While there is no indication that such an attack occurred at the GGD, it would also be impossible to detect one if it did occur. Even before the personal data breach in January 2021, it was already reported in the media that users had access to all information in the database, which would certainly get criminal entities interested. Combined with hiring practices that were focused on hiring as many people as possible instead of hiring people securely, with a VOG that was only required after six weeks, it isn’t unlikely that criminal actors would be able to access these systems.

5.6 Overview of Probable GDPR Violations

Based on the information from this section, I have identified the corresponding probable GDPR article violations. This section will provide an overview of these violations. Each subsection explains the definition of the article, why this investigation finds it probable that it may have been violated, and references findings to where they have been mentioned before in this section. The probable violations match the allegations related to the GDPR under the ICAM lawsuit, which is a violation of articles 5, 25, and 32 [161], in addition to articles 24, 33, 34, 35, 36, and 89.

5.6.1 Article 5 - Principles relating to the processing of personal data

Under article 5 of the GDPR, the processing of personal data must be done in accordance with various principles. Probable violates relate to subsections c and f, which state that personal data shall be "adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed" and "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorized or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organizational measures" [31, p. 35].

However, as this article is one of the basic principles of the GDPR under chapter 2, probable violations of this article are based on specific probable violations related to other articles of the GDPR. Therefore, any probable violation of this article will be determined by a probable violation in either articles 24, 25, and/or 32 of the GDPR, as discussed in this section.

5.6.2 Article 24 - Responsibility of the controller

Under Article 24 of the GDPR, the controller (in this case the GGD) "shall implement appropriate technical and organizational measures to ensure and to be able to demonstrate that processing is performed in accordance with this regulation" [31, p. 47]. A measure is a broad term as it refers to all actions that are appropriate to make the processing compliant with the GDPR, either through organizational or technical measures. For data processing to be secure, both organizational and technical measures must be taken although, in practice, there is not always a clear distinction between them as they can overlap [162].

The GDPR does not define what a technical measure is but does provide examples such as securing access to data, for example using password protection, and the transfer of data, for example using encryption [163]. Examples of organizational measures are data audits, activity logs, and internal training of employees by the Data Protection Officer [164]. Recital 78 of the GDPR specifies additional measures such as pseudonymizing personal data as soon as possible, ensuring transparency with regard to the functions and processing of personal data, enabling the data subject to monitor the data processing, and enabling the controller to create and improve security features [165]. This thesis will use measures beyond these examples, as a measure refers to all actions taken by the GGD.

Although CoronIT and HPZone (lite) were managed and further developed by the umbrella organization GGD GHOR Nederland, a significant number of organizations were involved in the processing of information. As stated in Section 5.5.3, data was processed by 25 different GGD organizations that each belong to a security region, GGD GHOR, six different national partner organizations (call centers and other external organizations), employment agencies, and IT suppliers. Although managing such a large number of organizations can already be problematic, this was further complicated by the fact that, as stated in Section 5.4.5, there was no clear overview of which organizations were in charge of which areas. As stated in Section 5.5.3, there was also no clear agreement related to various security aspects, which further increased the chance of new personal data breaches.

There were major problems related to securing access to data for internal employees, although improvements could also have been made to external security. However, as this study has not found any incidents related to data being compromised from the outside, this will not be comprehensively covered. As stated in Section 5.3.4, access to the various systems was achieved using an URL reachable from non-company-issued devices, without being logged in to a secure working environment, which presents a potential risk. Although this is mitigated by the requirement to use Two-Factor-Authentication. The lack of using company-issued devices also introduces various risks such as an inability to know if devices meet all required security measures, the possibility that non-updated devices are at risk of certain attacks, and that devices may already be compromised by malware. Combined with many employees working at home, out of view of any organization, this is a significant risk when working with sensitive information on a massive scale.

Data access for internal employees was managed differently between the two systems, although the GGDs approach to data resulted in the same problems occurring in both systems. Data access to HPZone (Lite) was divided into each security region, with data stored at each GGD, however, it was decided that employees could receive access to different regions. This process of giving access was not properly managed, which resulted in employees continuing to have access long after their access should have been removed. All employees also had access to all functionalities of HPZone, including an export functionality able to export significant amounts of entries without any kinds of restrictions or monitoring. CoronIT suffered from a similar design choice, where it was decided to work in a centralized system with all employees being able to fill in demand at every region and every job, instead of the siloed approach HPZone used before the Covid-19 pandemic. This was combined with a search functionality that was able to look up all individuals with minimal information and a manual logging functionality that stored logging information for just 1 month.

Organizational measures surrounding employees themselves were also inadequate, especially concerning the hiring and training process. As stated in Section 5.4.3, the hiring process was greatly simplified resulting in anyone being able to be hired in as little as a 2-minute phone call. To prevent malicious actors from finding employment at the GGD or any contracted organization, they made use of a ‘VOG’, although in a way that made this approach ineffective. Employees would have 6 weeks to hand in a VOG, while already being able to start working. In some instances, this step was even skipped entirely. As a VOG is a governmental declaration that detects if there are any objections to starting in a certain function, for example, if there is any problem with this individual working with sensitive medical information, this is an essential step that was constructed in a way that benefitted speed over security. Combined with the lack of an automatic logging system, this is a clear danger.

The training of employees, as stated in Section 5.4.4, was also inadequate. The systems, in general, were complicated, with employees stating that they found it difficult to work with, but would have to start working anyway. While internal GGD employees received training, the quality of training for external organizations was far below that of the GGD. External employees stated that they were violating the GDPR without even knowing what they did was wrong, as they never received GDPR-related training. Some training at external organizations was given by employees who had only a few days of experience themselves, which is far from the originally specified training for 3 days under medical professionals as stated as a requirement for working in HPZone. Employees would also go beyond protocol and access information they had access to, but would not be allowed to access based on their responsibilities. The lack of training also led to practices where employees would create WhatsApp groups where they would ask questions and receive help, which given the lack of GDPR-related training, led to employees sharing medical information within these groups with large numbers of other employees. The worst offenders of this practice would even share the contact information of famous individuals such as models, soccer players, celebrities, etc.

From these points, it is clear that the organizational and technical measures taken to ensure and be able to demonstrate that processing is performed in accordance with this regulation were likely to be inadequate for any system, certainly for a healthcare-related system containing detailed information of millions of individuals in the Netherlands.

5.6.3 Article 25 - Data protection by design and by default

Under article 25 of the GDPR, (in this case the GGD) “the controller shall, both at the time of the determination of the means for processing and at the time of the processing itself, implement appropriate technical and organizational measures” [31, p. 48], which “applies to the amount of personal data collected, the extent of their processing, the period of their storage and their accessibility” [31, p. 48]. As this article shares significant overlap with the previous article, this section will focus on the accessibility of the data, where the GGD has prioritized speed over security.

As stated in Section 5.3.2, HPZone and CoronIT had different approaches to data management. HPZone was initially federated, with each GGD being responsible for its own region, without any connections between the various regions. As HPZone was used by a small team of highly skilled medical professionals, employees would have full access to all data and functionalities in their region. These functionalities included functions able to export a significant amount of data outside of the system. The system also lacked a monitoring aspect, as determined by the “Axis into ICT” investigation stated in Section 5.2.2.

During the Covid-19 pandemic, a significant number of untrained employees would be added to these systems, however, GGD GHOR did not make any changes to the system to make it more secure now that the number of users had grown, with people not being medical professionals. Nor was the export functionality restricted to a limited number of users. Instead, GGD GHOR made the system more insecure by creating HPZone Lite, a version that was able to connect each region, so that access could be given to employees in one region, to support another. To make matters worse, there was no clear overview of access being given to employees, which resulted in many employees keeping their increased access, far beyond the need of their role. The deficiencies in HPZone proved to be so big that the only possible action was to replace it with GGD Contact.

CoronIT was developed as a centralized system, with each GGD organization or external organization working in the same system. This is in contrast to the way the GGD is designed to work, as they are a federated organization each in charge of their region, as stated in Section 4.4. However, for the sake of efficiency, it was decided that each individual would be able to call each region, instead of the security region they live in, to schedule either a test or vaccination. While this is easier for the individual and allows for each employee to be able to support anywhere in the Netherlands, this is a significantly riskier approach than the federated approach initially used by HPZone. Which could have been used to achieve the same ability to support regions, while not increasing the level of data access from a single region to the entire country.

While CoronIT could be used to support multiple roles, which would have allowed for a separation of data, instead the decision was made to give all employees access to the data needed for all roles. This aids in all employees being able to support all roles at any given time, although it also carries the increased risk of data that is not required to be accessible at all times. A more secure approach would have been to only give temporary access when required by demand.

5.6.4 Article 32 - Security of processing

Under article 32 of the GDPR, organizations are required to “implement appropriate technical and organizational measures to ensure a level of security appropriate to the risk” [31, pp. 51-52]. As this article shares significant overlap with both Article 24 and Article 25 which belong to the same chapter, a violation of this article is based on the violation of both previously stated articles.

5.6.5 Article 33 - Notification of a personal data breach to the supervisory authority

Under article 33 of the GDPR, in the case of a personal data breach, “the controller shall without undue delay and, where feasible, not later than 72 hours after having become aware of it, notify the personal data breach to the supervisory authority” [31, p. 52]. The concept of personal data breaches has already been described in Section 4.3.2, which in the context of this article refers to every incident identified in this thesis where data was leaked from either CoronIT or HPZone (Lite), including the smaller data leaks and the major data leak reported by RTL Nieuws.

Given that most personal data breaches are very small in scope, affecting only a single individual, it is important to specify which approach this thesis will use toward personal data breaches. As discussed in Section 4.3.2, most of these breaches are unrelated to the system used, stemming from incorrect information entered into the system or from information being delivered to the wrong recipient. Therefore, this article will only be addressed by examing larger personal data breaches. Primarily through employees accessing information from an individual beyond their duties, although it is also important to note the sharing of data in WhatsApp groups.

In every instance identified in this thesis, where the GGD became aware of a personal data breach, it appears that these personal data breaches have been reported to the supervisory authority, the AP. However, these incidents also share a common factor. In every single instance, the GGD was not aware of any such data breach occurring, until they were notified by an individual or organization, to point out this breach. As such, while the GGD adheres to the “after having become aware of it” aspect of this article, there is a potential problem.

According to the DPIA from CoronIT, the logging of employee activity is not activated in some instances, with no newer information able to verify that this has been activated. It further states that logging is not being actively periodically or automatically controlled, nor is there any kind of policy or agreement on who is responsible for executing these controls. It was recommended in this same document to ensure that logging is enabled and monitored, as otherwise, CoronIT systems would contain a high level of risk. The DPIA from HPZone Lite shows a similar problem, where it is not clear if each GGD is monitoring user activity internally, nor is it clear if this is done for external organizations. Based on the findings from 5.5.2, there is no evidence that this changed before the occurrence of the major personal data breach in January 2021. As such, the GGD made it nearly impossible to become aware of personal data breaches in the first place. Creating a situation where there is no awareness of data leaks in the first place can’t be in accordance with the spirit of the law.

As the GGD does not employ any automatic logging analysis, the GGD was not able to detect any, or at least the vast majority of, personal data breaches. Only after such incidents have been reported to the GGD or had reached the news were they able to take action. While it is likely that this covers all cases related to BNers and other known individuals, it is unlikely that this extends to the employees that illegally accessed the data from their friends and family. Which suggests a failure to report every instance to the AP. It is likely that if every one of these instances was reported to the AP, the AP would have reacted sooner, which would then have been reported in the news.

One of the bigger personal data breaches is related to the sharing of personal information in WhatsApp groups. Where employees would ask for help without censoring any personal information. As this is not strictly required for the processing of personal data, this would constitute a personal data breach. However, it is unclear whether this practice has been reported to the AP, nor that they would be able to identify each individual affected by this practice. While the various GGDs did aim to prevent the sharing of data in WhatsApp groups by outlawing the practice, the results were that these groups would then operate outside of the confines of the GGD. Making it even more difficult to both identify and report data breaches.

The major personal data breach was also correctly reported by the GGD, although there is a peculiarity in the way this report was handled. On the 22nd of January 2021, as discussed in Section 5.5.3, RTL Nieuws reported to GGD GHOR that a data breach of significant scope had occurred, with the possibility that data from tens of thousands of individuals had been stolen. GGD GHOR then informed the AP of this fact. Which is within the 72-hour time limit. However, on the 27th of January 2021, the AP immediately demanded clarification from the GGD about the personal data breach that was reported by RTL Nieuws on the 25th of January 2021 [166].

As the AP had already been informed about this personal data breach, there should have been no need to demand clarification. However, the AP's announcement makes it seem like they were unaware of any such report. This suggests that the AP is either not able to handle the report of a major personal data breach from a large governmental organization, or that GGD GHOR misrepresented the scope of the personal data breach.

The second of which would be a violation of Article 33, under paragraph 3, section a, which states that organizations are required to "describe the nature of the personal data breach including where possible, the categories and the approximate number of data subjects concerned and the categories and the approximate number of personal data records concerned". However, this cannot be determined without information about this report to the AP. It is also beyond the scope of this thesis as both GGD GHOR and the AP do not fall under the scope of the WOO.

5.6.6 Article 34 - Communication of a personal data breach to the data subject

Under article 34 of the GDPR, when a data breach occurs and the breach is "likely to result in a high risk to the rights and freedoms of natural persons", the controller is required to communicate this to the data subject without undue delay [31, p. 52]. In clear and plain language describing the nature of the personal data breach. The data contained in CoronIT and HPZone lite is highly confidential and sensitive data, which would meet the requirements set out in this article. Table 9 and Table 10 in Section 5.3.1 show numerous data fields that fall under a special category.

This article follows the same reasoning, from article 33, that creating a situation where there is no awareness of personal data breaches in the first place, cannot be in accordance with the spirit of the law. Without awareness of personal data breaches occurring, data subjects naturally can not be informed about any personal data breach.

While numerous personal data breaches occurred during the Covid-19 pandemic, this section will focus on two types of these personal data breaches. The first is the previously mentioned WhatsApp groups that led to medical information being shared in large groups of employees, the second is the major personal data breach that occurred in January 2021. Other data breaches are smaller in scope and impossible to investigate from an outside perspective.

As stated in Section 5.5, Whatsapp groups were used among employees to share data either maliciously or in an inherently insecure way to ask and receive help from other employees. While any incident where data would be shared in these groups would be a personal data breach that would be required to be reported by the AP, there is no indication that the AP knew about this practice nor was it ever covered in the news. As this practice was not common knowledge, it is unlikely that any data subject was contacted about this practice. Instead, it was employees that reported this practice to the news, although only at a later stage. Given that the information in these groups would have been easily identifiable, it would be relatively easy for the GGD to contact each data subject and notify them of a personal data breach.

With the major personal data breach, occurring in January of 2021, it is unlikely that every affected individual has been individually contacted. The GGD was unaware of any such data breach occurring and as a result, their logging information would not be able to cover the majority of the time that flaws in these systems were present, nor is it likely that all data breaches connected to this breach would have been detected. While 1.250 individuals were individually contacted, Section 5.5.3 suggests that the scope of the data breach is likely significantly higher. However, the GGD is likely not in breach of Article 34 over this instance as it is not a requirement for data subjects to be individually contacted. Instead, the GDPR requires organizations to notify individuals in general, such as with a post on the organizational website of a news report, which is the way GGD GHOR used [125].

5.6.7 Article 35 – Data protection impact assessment

Article 35 of the General Data Protection Regulation requires a Data Protection Impact Assessment (DPIA) to be conducted in instances where the processing of personal data “is likely to result in a high risk to the rights and freedoms of natural persons” [31, p. 54]. A DPIA shall be required in particular if one of the following applies:

- A systematic and extensive evaluation of personal aspects relating to natural persons which are based on automated processing, including profiling, and on which decisions are based that produce legal effects concerning the natural person or similarly significantly affect the natural person;
- Processing on a large scale of special categories of data referred to in Article 9(1), or of personal data relating to criminal convictions and offenses referred to in Article 10; or
- Systematic monitoring of a publicly accessible area on a large scale.

The systems used by the GGD meet all these criteria as they operate on sensitive data and have the ability to significantly affect individuals, they operate on medical data which is listed as part of article 9 of the GDPR ‘Processing of special categories of personal data’ and there was systemic monitoring of the entire nation. Table 9 and Table 10 in Section 5.3.1 depict the data that was processed using these systems and which data falls under these special categories.

As stated in Section 5.2.3.2, the Ministry of VWS believes that source and contacts platforms and by extension, any other system used to support the Covid-19 pandemic requires the creation of a DPIA. However, all systems covered in this case (CoronIT and HPZone (Lite)) have not followed this procedure correctly. The risks stated in the initial version of the DPIA of CoronIT as well as the risk stated in the improved version of HPZone Lite after the major data breach in January 2021 further support this fact.

A DPIA was initially completed for CoronIT, but it does not seem to fully reflect the system that the GGD ended up using. Given the fact that the purpose or nature of data collection and processing has changed, as the process of vaccinating individuals is different from the process of testing individuals, a new or updated version of the DPIA should have been made. As this was not done, this is a probable violation of Article 35.

A DPIA was not completed for the legacy system HPZone, which was created before the implementation of the GDPR. While a DPIA would not be required for a legacy system, the significant increase in the scope of the system would necessitate the creation of a DPIA. Before the Covid pandemic, source and contact tracing was a small-scale activity that targeted a limited number of individuals. However, during the Covid pandemic, source and contact tracing was applied on a previously unprecedented scale, targeting a significant extent of the Dutch population. A legal advisor attending the Justice and Security Data Conference also holds the opinion that a DPIA would have been required to be created [133]. As this was not done, this is a probable violation of Article 35.

A DPIA was also not completed for HPZone Lite. However, given that this is a newly developed system, although derived from HPZone, it would not fall under the legacy system exception. As such, it requires the creation of a DPIA, even if HPZone would not. A DPIA was later made for this system after the personal data breach of January 2021, which already took into required improvements into account. As this DPIA identified high risks in its current state, it would certainly warrant the creation of one at the start of creating HPZone Lite. As this was not done, this is a probable violation of Article 35.

GGD GHORs decisions to not create a DPIA for at least HPZone Lite may have been motivated by the recognition that the system would not meet modern security standards and could not address the deficiencies of its predecessor, HPZone. A later investigation into HPZone (Lite) after the personal data breach concluded that it did not meet many legal requirements, such as the NEN7510, NEN7512, ISO27001, and the GDPR [124, p. 29]. As the system was found to be unsuitable for further use, even outside of a pandemic, it led to the decision to completely replace this system with GGD Contact.

The DPIA for CoronIT did not accurately reflect the system but did contain ten different areas that were deemed at high risk [124, pp. 294-302]. With mitigation methods, these risks would be reduced to medium or low levels. However, the problem is that these mitigation methods, at least concerning logging [124, pp. 295-296], were not implemented at all (at least before the personal data breach).

In an interview with a governmental Data Protection Officer [89], he stated that an official admission of the system's insecurity would have significant implications for public health. It is possible that a significant number of individuals would refuse to use this and any future systems, hindering their ability to receive testing and vaccination. This could explain why no DPIA was made initially for HPZone (Lite) and why the risks identified in CoronIT were all reduced to medium or lower. It does not explain why GGD GHOR would not implement the mitigation methods they stated, however. While public health may have played a role in these decisions, creating a DPIA is still a legal requirement that cannot be avoided with data this sensitive and at this scale. As it can potentially put the sensitive medical data of millions of citizens at risk.

5.6.8 Article 36 - Prior consultation

Article 36 of the General Data Protection Regulation requires controllers, such as the GGD, to consult with the supervisory authority (in this case, the AP) before processing may begin if the Data Protection Impact Assessment under article 35 indicates that the processing would result in a high risk without mitigating measures. Given the sensitive nature of the data being processed on a national scale and the fact that the corona systems met all three criteria for conducting a DPIA, it is likely that the GGD would have been required to submit their DPIA for approval. However, no such submission was made for any of the three systems. And given that societal importance can also be a reason to submit a DPIA according to the DPO of the Corona Melder App [136, p. 2], all systems could be regarded as meeting this requirement.

The problematic aspect of this is that for HPZone and HPZone Lite, no DPIA was made in the first place. As such, there is no DPIA to indicate that there are significant risks nor is it possible to determine if the countermeasures are enough to mitigate these same risks. Therefore, even though the system might not be safe, a situation has been created where the AP does not have to be consulted. As these systems, as stated in the previous section, did not meet many legal standards, including compliance with the GDPR, it is unlikely that such a document would have indicated that there were no risks. The later DPIA of HPZone Lite, which was created after January 2021, about the current status of the system does prove that such as report should have been made as it identified six areas with a high level of risk to the data subject.

A DPIA was created for CoronIT, however, this document does not accurately reflect the way data was collected and processed as it was not updated since the very start of development. However, in this document, many areas were deemed at high risk [124, pp. 294-302]. With mitigation methods, these risks could be reduced to medium or low levels, which would not require a submission to the AP. However, as further versions of the system did not include these mitigation methods, any updated version of a DPIA would have to conclude that these areas remained a high risk. Which would require consulting the AP.

Given the nature of the data processed by the GGD, the massive number of employees working at home that received little training, and the GGDs own opinion that at least HPZone would not be able to meet current standards, it is unlikely that any DPIA could conclude that there weren't any risks. As the AP's investigation and reporting, a year after the systems were improved, still deemed these systems to be unsafe, it is unlikely that CoronIT or DPIA would have passed the evaluation of their DPIA, requiring a system redesign.

5.6.9 Article 89 - Safeguards and derogations relating to processing for archiving purposes in the public interest, scientific or historical research purposes, or statistical purposes

Article 89 of the GDPR specifies the conditions under which data can be used beyond the purpose of its collection if it is in the interest of "archiving purposes in the public interest, scientific or historical research purposes or statistical purposes". Provided this is done under appropriate safeguards. Including ensuring that technical and organizational measures are in place to fulfill the principle of data minimization.

This article has been included not due to a potential breach of this article where data was shared without the appropriate safeguards, but because organizations are unwilling to make use of this exception. Section 2.1.3 describes that data from the GGD was shared to do a study into the excess mortality rate due to the Covid-19 pandemic, yet this was not provided to all parties. Instead, this data was only shared with the RIVM which was already a recipient of GGD data, as well as the CBS, a trusted third-party organization. This goes against the request by the House of Representatives to allow academic parties to an independent investigation

The reasoning provided by the GGD is that they are hesitant to provide this access as they believe it would constitute a violation of the GDPR. This is in contrast to both this article and the fact that data was shared with these other organizations. The FAIR architecture, specified in Section 0, will aim to address this.

6. Governmental personnel's understanding of the reason problems occur in governmental systems

This section aims to explore the perspectives of governmental personnel regarding the reasons behind problems occurring in governmental systems. The investigation involved attending two major governmental conferences, which drew significant participation from personnel representing various organizations. The organization of such conferences by the Dutch government on topics like data management and ICT usage indicates the importance of these areas and the ongoing need for improvement.

While healthcare organizations were not present at the conferences, it is crucial to emphasize the application of privacy by design and by default, as well as data minimization, in all systems processing personal data that poses risks to data subjects. Therefore, knowledge related to protecting valuable data in other organizations can also be applicable to healthcare-related organizations.

On September 14, 2022, I attended a government-organized conference that brought together high-level governmental employees and representatives from the private sector to discuss digitalization. The conference had a significant turnout, with over 450 individuals in attendance. During the conference, I actively participated in panels and conducted informal interviews with attendees from numerous organizations.

On October 18, 2022, I attended a conference organized by the Ministry of Justice and Security, titled "Data, in the Service of a Just Society." This conference included participation from various departments within the ministry, such as the national police, the National Coordinator for Security and Counterterrorism, the Netherlands Forensic Institute, the department responsible for investigating cybercrime, the Custodial Institutions Agency, the Inspectorate of Justice and Security, and high-level government employees from the Ministry of Justice and Security.

The findings from both conferences have been combined and categorized into the following sections: data security, which examines differences in data security standards and data processing methods across various organizations; perspectives on supervisory organizations; an evaluation of general ICT problems related to the parliamentary inquiry described in Section 4.1; and the lack of a common language and control over data, with relation to the GDPR. It is worth noting that interviewees, in general, displayed limited familiarity with FAIR. However, this is not surprising, as the Dutch government is still exploring the potential value of applying FAIR in organizations, particularly in healthcare organizations.

6.1 Data Security

Data security holds immense importance for organizations, as highlighted by Ronald Damhof, Chief Data Officer of the Ministry of Justice and Security, who emphasizes it as one of the most crucial investments [167]. However, the approaches taken by organizations to improve data security often vary significantly. From implementing diverse schemes to segment data access to employing different security measures to prevent unauthorized breaches, each organization adopts its unique strategy. This observation is further supported by the fact that the organizations attending iBestuur and the data conference organized by the Ministry of Justice and Security had distinct approaches to improving data security. Even when organizations fall under the same larger organization, such as the organizations organized under the Ministry of Justice and Security, differences remained. An overview of these approaches can be found in Section 6.1.4.

While organizational differences exist, data security among organizations is often applied through standardized frameworks and shares common factors. An example of a standardized framework is the “Baseline informatiebeveiliging Overheid” framework (hence “BIO”) and an example of one of these common factors is the logging of user activity. It is also important to find a balance between usability and security, as the inclusion of too many security measures may prevent a system from being effective for the purpose it was designed for. Security remains a critical factor however, as while a system that lacks security measures may be less complex to create, and allow for data to be processed more efficiently, it may also prevent such a system from being allowed to be used due to not adhering to legal requirements such as the GDPR and relevant national data laws. Statements made in Sections 5.4.3 and 5.4.5, indicating that speed and efficiency were favored over data security, are contractionary to this balance and should prevent such a system from being allowed to be taken into active use.

6.1.1 Baseline informatiebeveiliging Overheid

The BIO is the most recent governmental information standard, with the implementation of these norms starting on the first on January 2019. This framework aimed to standardize information security across all levels of government, replacing the previous approach where each layer of government had its own framework, as well as to mandate previously established security norms. These layers of government are the national government, municipalities, provinces, and regional water authorities [168], which previously made use of the BIR [169], BIG [170], IBI [171], and BIWA [172], respectively.

It is important to note that these layers of government do not include the GGD or the security regions they are a part of. Therefore, the new BIO does not apply to these organizations. The security regions themselves aimed to fully comply with this standard by January 1, 2023, with all security regions making progress towards this goal, but not all security regions have been able to achieve this at this date [173]. The GGD decided not to participate with this standard, instead aiming to comply with the NEN7510 norm. Nonetheless, there is some level of cooperation between GGDs and the security regions on certain aspects of the BIO, such as awareness campaigns [174].

An evaluation conducted in December 2022 concluded that while the BIO has been successful, it has not fully ensured data security [175]. The positive aspects of the BIO include its wide acceptance across the entire government, serving as a basis for optimizing information security, providing a common language to strengthen secure cooperation within the government, and fostering trust within and between governments [175, p. 25]. However, the problem that persists is that hat the BIO can only lead to improvements if sufficient attention is given to preconditions, such as embedded risk management. Every organization needs to recognize the importance of information security and integrate it into its safety or risk culture, starting from the executive layer and extending throughout the organization [175, p. 27].

6.1.2 Monitoring User Activity

Monitoring user activity is a common practice among organizations, although the specifics may vary depending on the nature of the data being processed and the individuals involved [167]. In Section 6.1.4, we will delve into the differences observed across organizations in more detail. For this subsection, our focus is specifically on the healthcare sector, as monitoring user activity is a requirement under the NEN 7513 norm [176].

According to this norm, every interaction with a hospital system, for instance, must be logged. The purpose is to prevent unauthorized access to patients' records by medical professionals without a valid reason. A notable incident occurred in 2018 at the HagaZiekenhuis in The Hague, where numerous employees accessed the same patient dossier without any justification [177]. This breach was identified during a routine check of the monitoring logs, with no indication of an automatic system or user activity flagging. Considering that all 4,000 employees had unrestricted access to patient records across all hospital locations [178], this presented a significant risk.

To address this issue, some hospitals maintain a "people of interest" list, which includes Dutch celebrities and other prominent individuals. If any employee accesses the patient records of these individuals, their account is flagged for further investigation. Illegitimate searches lead to severe consequences, including termination. However, to minimize mistakes, the system prompts users to confirm their searches [167]

Some hospitals address this using a list of “people of interest” consisting of Dutch celebrities and other important individuals. If any employee accesses any of these patient records, their account is flagged for further investigation. If the search is determined to be illegitimate, severe action, including the termination of the individual, is taken. However, to prevent mistakes, the system asks you to confirm your search [167].

Given that this exact scenario occurred at the GGD, as stated in Section 5.4.4, and given that these systems are already in use by the healthcare sector, it is unclear why the GGD would not have implemented either these or similar measures to prevent this behavior from occurring. Even if the GGD would have believed such a system would not be required as they would not be processing data as sensitive as in a hospital, accessing private information from people of interest without justification is a common occurrence in general, beyond the healthcare sector [179].

6.1.3 A balance between usability and data security

While data security is an important aspect of any system, it should be balanced between security and usability, based on the impact of the data and the importance of the system [167]. Systems could be constructed in a manner where employees would need to request approval for any interaction, which would be incredibly secure, while at the same time making the system useless for any purpose. It’s also impossible to create a system that can 100% be guaranteed to be safe, as this would require not just preparing the system for any existing threat but anticipating future ones as well. Therefore, data security measures should be based on reaching a certain level of security, while trying to minimize the impact on the efficiency of the system.

6.1.4 Overview of Data Security standards in various organizations

This section contains an overview of the data security standards of the Custodial Institutions Agency, the Dutch institute for vulnerability disclosure, the Inspectorate of Justice and Security, the National Coordinator for Security and Counterterrorism, and the Netherlands Forensic Institute. All organizations place a strong emphasis on protecting data from outside attacks while varying in the level of internal security measures. Which are based on the size of the teams processing personal data, their level of experience, and the sensitivity of the data itself.

In general, if data is processed in teams of a limited size with experienced individuals, full access is given to employees processing that data, regardless of the sensitivity of the data. However, if either the size or quality of these teams changes, additional security measures are taken and employees will have their level of access strictly reduced. When working with particularly sensitive data, or if any breach has significant consequences, additional security measures will be taken, even if operating in small teams of specialized individuals.

The Custodial Institutions Agency, as described by an attendee from this organization, faces unique challenges as certain information breaches are not only a violation of the GDPR but may also lead to judicial or political problems [180]. For example, the recording of a conversation between a prisoner and their attorney would be cause for a judicial investigation, which could endanger a potential judicial proceeding and/or require a defense in parliament. Beyond this, there is no further information about the teams processing data in this organization, the data is highly sensitive. Although the uniqueness of this organization may lead to data security measures not being able to be incorporated into any other system.

The Dutch institute for vulnerability disclosure, as described by an attendee from this organization, takes certain security measures that were not specified in any other organizations discussed here [181]. Access to data in this organization is strictly divided, with data fields not being available to lower-level employees. Which matches the more privacy-oriented system that should have been used by the GGD. Only by special request to the highest-level individual in that organization, the director, may data be accessed on a higher level or may it be exported. As a further security measure, this information is stored on a physical drive that is not accessible from the internet.

The Inspectorate of Justice and Security monitors the data security standards of organizations under the umbrella of the Ministry of Justice and Security. Which also included matters related to cybersecurity and cyber threats. While this attendee was not able to disclose any information related to how data is processed in this organization or state how these inspections are executed, he was able to provide a general view of how these organizations protect their data and which security measures are taken internally [182].

In these organizations, data minimalization, as specified in the GDPR, is generally applied. A DPIA is also created before the processing of data. Organizations generally keep a record of user activity, although the exact way this is logged is confidential. However, in the past, some organizations have made it impossible to complete an investigation due to a lack of logging data. This is similar to the investigation into the personal data breach at the GGD, which as discussed in Section 5.5.3 would lack most data needed to fully investigate the scope of the personal data breach.

Data security at the National Coordinator for Counterterrorism and Security, as described by an attendee from this organization, focuses on both internal and external security [183]. One of the responsibilities of this organization is creating and communicating the routes prisoners take when they travel to different locations. As any mistake or data breach risks the escape of a prisoner, endangering society as a whole, access to data is strictly controlled, with all user activity being monitored. The data in this system is segmented and inaccessible to everyone that does not require access.

Data security at the Netherlands Forensic Institute, as described by an attendee from this organization, resembles a ‘data moat’ [184]. Data is well protected from hacking attempts from the outside world, but within the organization, no specific data protection methods are used. Each investigator employed at this organization only has access to the case they are assigned to investigate, however in the context of that case, they have access to all data, and their activity is not monitored nor logged [185].

Access to this case is provided based on Privileged Identity, which is a time-based authentication method that can provide access for segments of data. Employees must authenticate themselves using 2FA, upon which they gain access to the cases they are working on. As the authentication is time-based, at the end of a case or a certain amount of time has passed, they lose access to that case. As this was done manually in the GGD case, with numerous mistakes, this is a simple improvement that can reduce the risk of data breaches.

6.2 Supervisory organizations

Given that the supervisory organizations, the AcICT (Section 4.2) and the AP (Section 4.3), did not play any role during the GGD case as discussed in Section 5.2.3, these conferences were used to establish if this is a more general occurrence. The findings from this section generally come from the iBestuur conference as the various organizations from the Ministry of Justice and Security were unable to comment. However, attendees in general were limited in their responses or made statements that did not reflect governmental documentation.

When proposing the question if these supervisory organizations have enough resources to be able to carry out their mandated functions in the panel 'Wat is het probleem van Big Tech?' [186], the panel did not have a clear answer, instead stating that more research would be required by the government to establish this. The goal of this investigation would be to determine which tasks the organizations would not be able to perform due to a possible lack of funding, which could then be incorporated into the next coalition agreement. A representative from the department general of digitalization also stated that additional research would be required to determine if these supervisory organizations would need an increase in their budget [187]. However, as stated in Section 4.3 such an evaluation has already been conducted for the AP, which concluded that the AP would require a significantly increased budget to be able to handle its current responsibilities. The coalition agreement for 2023 does include an increase in the budget of the AP, however, the increase is less than this evaluation suggested. Such an evaluation has not been conducted for the AcICT, nor was their budget increased in the most recent coalition agreement.

While the AcICT, as stated in Section 4.2.2, is unable able to process the current number of projects that it is legally mandated to evaluate, the Secretary of State of digitalization, Alexandra van Huffelen, does not believe this to be related to the level of funding [77]. In her opinion, a sample-based approach would allow all projects to benefit from the results of published reports, which negates the requirement to evaluate all projects. Thereby improving the development of ICT projects in government in general. She also stated in her contact with the AcICT, no request for additional funding had been made, which would indicate that the AcICT does not have a lack of funding. However, this sample-based approach prevents the AcICT from being able to fulfill its second primary aim, which is to provide advice on changes that could increase the success and effectiveness of a project and have the authority to decline a project if it is deemed unnecessary or unlikely to succeed.

Determining the exact reason, whether it is a budgetary or organizational reason, is ultimately beyond the scope of this thesis, nor does it address the actual problem related to these organizations. The actual problem is that measures that were created to improve the success of major ICT projects exist but are not utilized, as stated in Section 5.2.3. It is likely that If these organizations were involved at any point before the major data breach in January 2021, some of the problematic elements could have been addressed. Which would have required either the submission of a DPIA to the AP or a submission of the project to the AcICT.

6.3 Reference to the Parliamentary Inquiry

The parliamentary inquiry of 2014, as discussed in Section 4.1, is an important benchmark for the general state of ICT development in the Netherlands. As such, attendees were asked to comment on this investigation and describe the improvements that had been made since this benchmark if there are any.

According to the secretary of state of digitalization, Alexandra van Huffelen, while many areas have seen improvements compared to this benchmark, there is still much to improve [77]. Communication in and between organizations has seen the biggest improvement, as well as the successful creation of the supervisory agency for projects, the AcICT. The management of governmental projects still needs improvement, however, as many projects still exceed their budget and/or schedule. Many governmental projects are still unable to achieve what they set out to do or achieve these with serious problems being present in the design and implementation of systems. There has also been little improvement in the government's reliance on ICT contractors, as there is still a lack of internal ICT knowledge in governmental organizations.

According to an attendee from the policy division of the Ministry of Justice and Security [188], many problems that the government aims to solve with ICT might be policy and management related. As such, creating an ICT project with an ICT solution is the wrong approach towards a problem that can be solved at either a different level or via a different method. This appears similar to one of the findings from the inquiry that the government has significant enthusiasm to solve problems with ICT, leading to ICT being used in projects that don’t require it, with promises being made that are infeasible or impossible on a technical level.

6.4 Challenges in Establishing a Common Data Language and Ensuring Data Control

During a panel discussion involving high-ranking governmental personnel and academic professors, it was acknowledged that a significant obstacle to the processing and sharing of data is the lack of a common language [189]. A challenge that is not limited to governmental systems, but extends to various sectors, including the healthcare sector as a whole. A report by the RIVM on this challenge emphasized the critical need for a common language in healthcare, given the increasing exchange of health information between healthcare providers and patients through information systems. To ensure efficient and safe healthcare delivery, the shared information must have consistent meaning and purpose. Additionally, agreements that define the usage of technical terms in computer systems are essential for effective information exchange and reuse [190, p. 3].

In addition to the terminology and format barrier, data processing must adhere to relevant legislation and public policy, with the General Data Protection Regulation (GDPR) serving as the primary legal framework. The panel acknowledges a lack of data control within organizations as a prevalent issue, resulting in organizations being unable to properly control the data processing process. As a result, data is often shared without a solid basis. In other instances, organizations may be hesitant to share their data at all, due to the previously stated oversharing of data. Even approved investigations, may be unable to receive data from other organizations, as identified in Section 2.1.3.

To address these challenges, organizations should establish a comprehensive understanding of their data. This includes documenting informational metadata that provides context, outlines the relationships between various systems and data, and specifies the duration of data storage [167].

Furthermore, organizations should document the data's origin, format, and purpose. While a Data Protection Impact Assessment (DPIA) typically encompasses this information, many legacy systems do not have a DPIA, as the GDPR does not mandate one for unchanged data processing activities. Even when a DPIA is required, its creation is not consistently enforced, as indicated by the findings presented in Sections 5.2.3.1 and 5.6.7.

Another significant challenge relates to the individuals involved in data processing. Often, organizations do not allocate sufficient time to train employees in areas beyond their specialization. For example, data engineers would greatly benefit from having a general understanding of the legal aspects of data processing. Unfortunately, academic curricula rarely cover this interdisciplinary knowledge, as highlighted by the academic professors in the panel discussion. Consequently, employees excel in the technical aspects of their job but lack awareness of the legal implications of data processing. This knowledge gap poses a significant risk that organizations must monitor instead of addressing the problem at the root, preventing misconduct from occurring in the first place.

To overcome these challenges, the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, as described in Section 2.2, offer a potential solution. FAIR could act as the recommended common language specified by the RIVM, ensuring consistent understanding and usage of data across participating organizations. Moreover, FAIR empowers organizations to retain control over their data by providing access based on specific conditions, without resorting to file-sharing methods. The metadata component of FAIR enables a comprehensive description of the data, including context, system-data relationships, data storage duration, origin, format, and purpose.

Although FAIR does not explicitly include training, it can be complemented by targeted training programs. For example, the VODAN-Africa project incorporates training and certification to facilitate higher-level data access and enhance network contributions. Additionally, organizations can provide traditional training on the legal aspects of data processing to further support their workforce. Combining FAIR implementation with focused training on FAIR and legal considerations ensures a holistic approach to data management.

6.5 Conclusion

The understanding of governmental personnel regarding the issues in governmental systems aligns significantly with the problems identified in the parliamentary inquiry discussed in Section 4.1, which took place nine years ago. Despite the introduction of various improvements and initiatives to tackle these issues, their effectiveness is undermined when organizations are either unable or unwilling to implement them. For example, while the BIO effectively standardizes security practices and offers significant improvement in this area, it becomes ineffective when organizations are not mandated and/or willing to comply with these standards. Neglecting to adopt specific security measures implemented successfully by other organizations results in recurring problems that have already been addressed. The mere existence of supervisory organizations alone does not lead to improvements; active involvement in projects is necessary for project-specific enhancements. If these measures can be circumvented through various means, no improvement or initiative can effectively address the problems that arise in governmental systems.

Real improvement requires organizations to recognize the value of information security and embed it within their safety and risk culture, from the executive level down to the entire organization. Organizations must carefully balance system usability with data security, considering the sensitivity and scope of data processing. While some organizations have achieved this balance successfully, the GGD has not demonstrated this capability during the Covid-19 pandemic, primarily due to the lack of oversight.

Another crucial aspect lacking in organizations is a common language that ensures shared data has consistent meaning across different entities. This is especially critical in the healthcare sector to ensure efficient and secure care provision. The common language should also include agreements defining the usage of technical terms in computer systems for information exchange and reuse. Although data must be processed in accordance with legislation and public policy, such as the General Data Protection Regulation (GDPR), organizations currently lack control over their data, resulting in data sharing without a solid foundation.

The core technology of FAIR can address the technical problems stated in this section, but without organizational measures, it is unlikely to be able to address the problems that occur in governmental healthcare systems. It is important that unlikely previous improvements and initiatives, this initiative will not only be created but applied to organizations. Which may require organizations to realize the value of information security. Only if these requirements are met will any FAIR-based solution be able to have any effect.

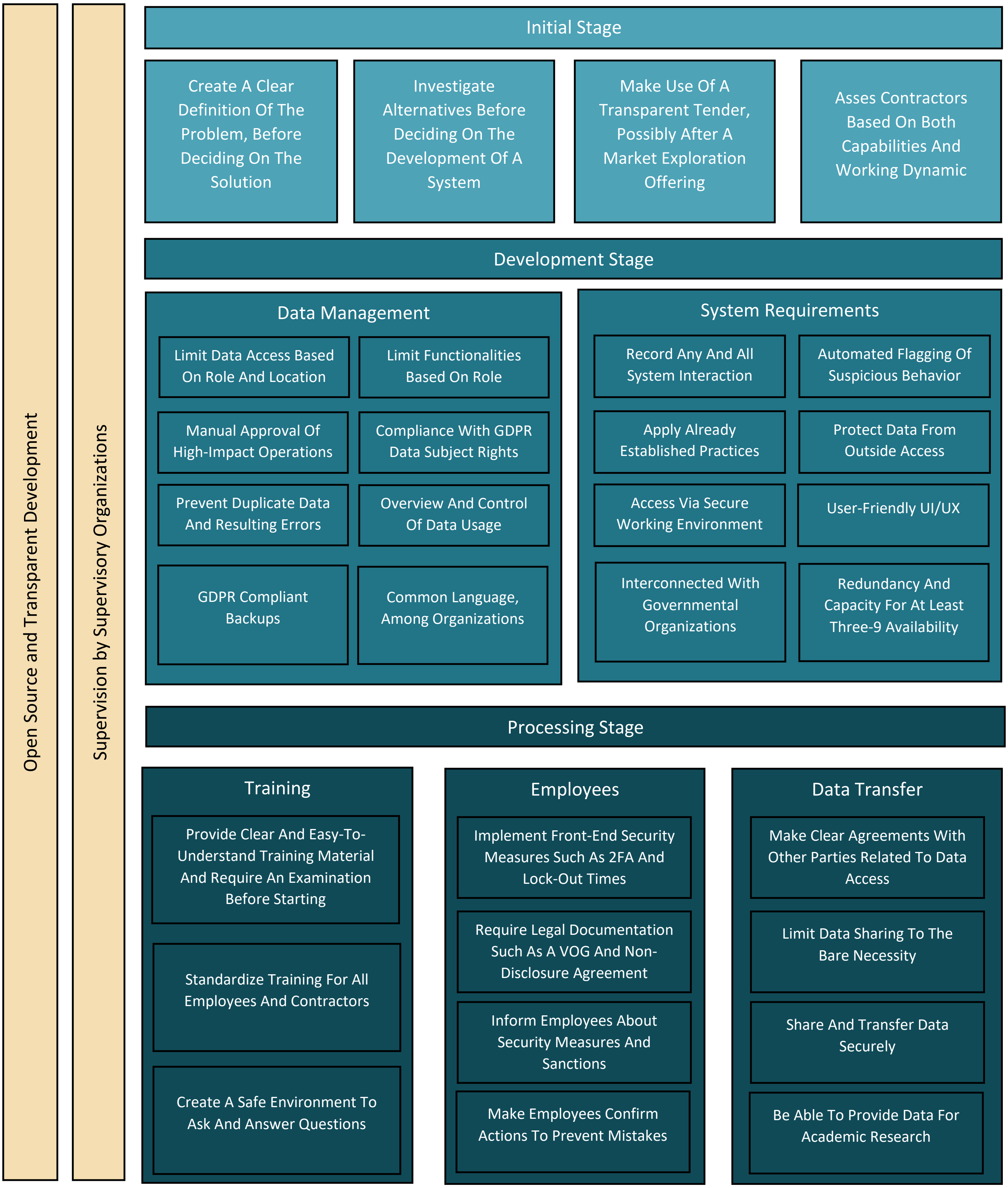
7. Specification and Requirements of a GDPR-compliant system

This section contains the specifications and requirements for a GDPR-compliant healthcare system, with considerations being made to ensure that the system is not only secure but also usable. In addition to this, the specifications and requirements are based on a healthcare ICT system, used by a significant amount of people, processing sensitive data and requiring significant organizational and technical measures to achieve GPDR compliance. It's also important to take into consideration existing improvements, initiatives, and best practices instead of starting entirely from scratch, as no project is ever truly unique in its design or purpose and lessons should be taken from other projects and sectors.

This has been based on the general requirements of the GDPR, the problems encountered during the case study investigation as discussed in Section 5, a more general investigation into the problems occurring in governmental systems as discussed in Section 6, and specifications and requirements based on the experience of Mustafa Kedilioglu, who was a project developer at the Corona Melder app, and of Ron Roozendaal, the former CIO of the Ministry of VWS and the former deputy general of Digitalization.

The specifications and requirements have been divided into four different segments, which are related to the overall project, the initial stage of the project, the development stage of the project, and the processing stage of the project. An overview of this has been depicted in Figure 9, with Sections 7.1 to 7.4 discussing the various elements in detail.

For use in the FAIR-based architecture in Section 8, these specifications and requirements have been evaluated based on if they are either core principles of FAIR, can be achieved with FAIR although it depends on its implementation, have significant value as an extension to FAIR and have therefore been included or have not been included in the FAIR-based architecture. An overview of this has been depicted in Figure 10, with Sections 7.5.1 to 7.5.4 discussing the various elements in detail.



Open Source and Transparent Development

Supervision by Supervisory Organizations

Initial Stage

Create A Clear Definition Of The Problem, Before Deciding On The Solution

Investigate Alternatives Before Deciding On The Development Of A System

Make Use Of A Transparent Tender, Possibly After A Market Exploration Offering

Asses Contractors Based On Both Capabilities And Working Dynamic

Development Stage

Data Management

Limit Data Access Based On Role And Location

Limit Functionalities Based On Role

Manual Approval Of High-Impact Operations

Compliance With GDPR Data Subject Rights

Prevent Duplicate Data And Resulting Errors

Overview And Control Of Data Usage

GDPR Compliant Backups

Common Language, Among Organizations

System Requirements

Record Any And All System Interaction

Automated Flagging Of Suspicious Behavior

Apply Already Established Practices

Protect Data From Outside Access

Access Via Secure Working Environment

User-Friendly UI/UX

Interconnected With Governmental Organizations

Redundancy And Capacity For At Least Three-9 Availability

Processing Stage

Training

Provide Clear And Easy-To-Understand Training Material And Require An Examination Before Starting

Standardize Training For All Employees And Contractors

Create A Safe Environment To Ask And Answer Questions

Employees

Implement Front-End Security Measures Such As 2FA And Lock-Out Times

Require Legal Documentation Such As A VOG And Non-Disclosure Agreement

Inform Employees About Security Measures And Sanctions

Make Employees Confirm Actions To Prevent Mistakes

Data Transfer

Make Clear Agreements With Other Parties Related To Data Access

Limit Data Sharing To The Bare Necessity

Share And Transfer Data Securely

Be Able To Provide Data For Academic Research

Figure 9 - High-level overview of Specifications and Requirements

7.1 Specification and requirements related to the overall project

The aspects of open source and transparent development and supervision by supervisory organizations apply to the entire duration of the project, acting as both a way to ensure that technical and organizational measures are implemented and to ensure that people perceive the system as secure. This second point is especially important as it must be prevented at all costs that individuals would be hesitant of interacting with the healthcare system, thereby leading to actual harm to public health. Both aspects also share significant overlap, although the first aspect is focusses on making development transparent to the public, and the second focuses on making development transparent to governmental organizations, whose reports are often accessible to the public.

7.1.1 Open Source and Transparent development

One of the findings from the Literature review, as stated in Section 1.1, is that it is not enough to ensure that systems are secure and can keep information private. Individuals also need to perceive it, which can be unrelated to the actual aspects of the system. If this is not achieved, or worse if personal data breaches occur that may make people hesitant to interact with the healthcare sector, then this has serious negative consequences for public health. Therefore, transparency throughout any project, in combination with open-source development, further increases transparency and may also result in systems becoming more secure.

During the development of the Corona Melder App, this approach was seen as critical to its success, as stated in Section 4.6, as its success depended entirely on the adoption by the population. Without enough people, an application that registers if you have been in contact with someone that has corona would be useless, as many people that have corona wouldn't use the app and would therefore not be detected. At the same time, healthy people that don't have the app also wouldn't be informed if they came across someone using the app that was marked as infected. In this project, the Appathon had the dual purpose of being "intended to gain the trust of the people" and "poll how open people are to the idea of using these kinds of apps in the first place" [191].

A later evaluation by the BIT of the process of development and implementation of the Corona Melder App determined that this project is too unique to be used as a blueprint for other ICT projects, however, this does not appear to be fully accurate. According to Ron Roozendaal, "while it is true that this same development process can't be applied to every project, it does not mean that it can't be applied to many of them" [115].

While this will require an organizational change, this approach has already started to be applied, with for example parts of DigID being made open source [192] and the intention to do this for future versions as well [193]. There is also a Ministerial order that requires at least the login component of systems to be made open source, so that "it is known to everyone how to login functions work, and it is therefore verifiable how the processing of personal data takes place" [194, p. 4]. The 'algoritme register' [195] is another example of an initiative that aims to improve transparency, by providing insight into which organizations utilize algorithms for specific purposes, and by detailing which data is used to make a decision [196]. With even the register itself being open source [197]. It should be noted, however, that these initiatives are all very recent, with the Corona Melder App being the earliest example, from 2020, the algorithm register is from the 21 of December of 2022, and DigID has been made partially open source in January of 2023.

In addition to improving transparency, an open-source approach allows for the use of the 'Wisdom of the crowd' effect, which was already successfully applied in the Corona Melder App Development

Process. While one person may be wrong, increasing the number of people increases accuracy up until the answer is close to right. An important caveat is that while this works for general knowledge, it likely doesn't apply to highly specialized knowledge where there is no reasonable expectation of the crowd knowing the right answers. As an open-source approach would likely only attract people with knowledge of software, it would self-select to a crowd that does have this experience, however.

Mustafa stated about this project that using an open-source approach and using open-source solutions or providing source code enables people to "think along, contribute to the solution and trace things that are not right in the software" and that "through more people looking at the code, it becomes more likely that any overlooked mistakes of unintended effects will be traced, likely in advance of any other type of analysis" [191]. This is why this approach has been made a requirement for the successful creation of such a healthcare system.

Transparency could also be expanded to include documents such as the DPIAs, which if a system is safe enough to use, would not be able to expose any design flaws that could be exploited. As if there were any, then it is unlikely that any such system would be safe enough to use. If information cannot be released, however, it should be released with some redacted information. With the critical note that redacted information should be kept to a minimum. In my investigation into the GGD case, I found two contrasting publications related to CoronIT based on the same WOO request, with the first published by GGD Zeeland [91] heavily redacting information while the same publication by GGD Hart voor Brabant [124] not redacted to this same degree, with information related to the risks to the data subject remaining visible. Therefore, as much information as possible should be released, and redacting information should be done with care, as it appears that there is no reason to redact information to the extent of the first publication.

7.1.2 Supervision by supervisory organizations

For this requirement, the most important organizations are the AcICT (Section 4.2) and the AP (Section 4.3). The role of the AcICT is to monitor the actual development process, to be able to detect and address flaws and mistakes in the implemented architecture, and to be able to suggest detailed improvements leading to a project having the highest level of success. The role of the AP is to ensure that the project is compliant with the GDPR, by analyzing the continuously updated DPIA and monitoring that the specified improvements have been implemented correctly in the system.

The AcICT needs to be involved to give project-specific advice, based on their second primary objective. Given the importance of any healthcare-related project, certainly, at this scale, its success is critical. As the AcICT is a highly skilled organization, with significant experience related to healthcare systems based on their evaluation of the replacement of Praeventus, they will likely be able to increase the chance of the success of a project. Although no supervisory organization can guarantee it, given that they aren't the ones implementing the project.

The AP is involved by law if a DPIA indicated there to be any high risk that cannot be mitigated, as discussed in Section 5.6.8. However, as this process can be circumvented by never creating or updating the DPIA for the project, as discussed in Section 5.6.7, this may not be enough to ensure that processes are reported. Therefore, any healthcare project should be required to both create a DPIA and submit it to the AP, by default. This has also been the conclusion of the governmental DPO on the Corona Melder App [136, p. 2], as discussed in Section 5.6.8, based on the fact that the societal importance of the system would require its submission to the AP.

7.2 Specifications and requirements related to the initial stage

The four specifications and requirements related to the initial stage are requirements that must be met before starting any project and result in any project having the highest chance of success. All of these could also be made public, in support of the transparent development of projects.

As stated in the parliamentary inquiry discussed in Section 4.1, a policy advisor as discussed in Section 6.3, and the statement from Mustafa, it is important to first establish what the actual problem that needs to be resolved is. The government has an unbridled enthusiasm for ICT projects, which is likely to result in the creation of one, even when other solutions may be better. Therefore, this is an essential step before any project can be started.

Before starting any project, alternatives need to be investigated. As there may be alternatives available to use right now, or with minor changes, that meet the requirements of the to-be-started project. If this is the case, then it is likely unwise to start the development of an entirely new project where success isn't guaranteed. In the GGD case, as discussed in Section 5, there were alternatives for both CoronIT and HPZone (Lite), such as Go.data and Praeventis, however, these were not seriously considered. While time and organizational constraints may ultimately prevent the use of alternatives, it at least warrants a serious investigation. This needs to be compared not only to the capabilities of any new project but also to its likelihood of success and development time.

Making use of a transparent market tender, published at TenderNed, is an important requirement that is also connected to the previously stated transparency aspect that should be applied to the entire project. It is also a legal requirement for governmental organizations under the "aanbestedingswet 2012" [114]. These market tenders allow individuals to request information from the government, provided that it is not confidential business information, such as which parties responded to the tender, which party was selected, and why. To show that governments do not favor any party and provide everyone an equal opportunity to compete in offering something that meets the requirements [191].

However, as stated in Section 5.2.1, no such tender had been published at TenderNed for CoronIT, based on article 2.32 paragraph 1 subsection 1, which exempts public tenders from having to be created in a crisis situation [114]. As no tender exists for HPZone Lite, it can be assumed that his exception was also used for the development of this system. Given that the Corona Melder App, as stated in Section 4.6, was developed during the same period yet did not use this exemption, it appears that a crisis is not a hindrance to this requirement. Therefore, any project should always create a tender, published on TenderNed, to ensure that this process is as fair and transparent as possible. Even if such a tender is made, the Corona Melder App also shows that it can still be decided to internally develop a project if there are no offers that meet the initial requirements. Which further reduces the use of such an exemption.

The previous statement that all parties should receive an equal opportunity does not result in every party being equally likely to be made responsible for the development of a project, however. When selecting a party to be responsible for the development of a project, it is important to select this party based on both their level of capability as well as their dynamics, as it is important that they can keep up with the dynamics of the government [191]. Any party should be examined based on the size of the organization, their level of related experience, their reputation, and the level of success of previous governmental projects. In the case study into the GGD, as stated in Section 5.2.1, this requirement was not followed and likely led to various problems. This requirement would likely have detected that while GGD GHOR may have been the logical choice as it is responsible for managing public health in the various security regions, the limited size of the organization, its lack of experience in both working in any comparable system and developing systems at a larger scale, the failure of developing new Praeventis at a much larger and more experiences organization and the already existing problems in HPZone negate this fact.

7.3 Specifications and requirements related to the development stage

This section describes the organizations and technical measures required for secure processing and data minimalization privacy by design and by default concerning the development of the system, which has been depicted in Figure 9, In further detail. This section has been divided into specifications and requirements related to the actual data management, and the requirements of the system. Although it could be argued that some of these aspects could be covered under both categories, for example, the requirement for GDPR-compliant back-ups.

7.3.1 Data Management

One of the most important requirements to ensure that data remains private and secure is to limit access to data as much as possible, concerning the role of the user that requires access and the location the user is employed at. Systems can be used to support multiple roles, therefore access should be designed in such a way as to ensure that users only have access to the data they need for that specific role. Given the federated nature of the healthcare sector, either with employees employed at a specific healthcare location or as part of the GGD in a specific security region, access should ideally be limited to that specific location. Access beyond the confines of these requirements may improve efficiency, but it would have significant implications for the security of the data and should be avoided whenever possible.

This does not mean that no access can be given beyond either a specific role or beyond a specific location, but it should be both temporary and require a unique login so that the data able to be accessed from a single instance is limited. The GGD case provides an example of why this requirement is important and the dangers of implementing it incorrectly. As stated in Section 5.3.2, data access in HPZone was limited to each security region, which is an inherently more secure approach. Data access was then expanded based on a system that allows access to data from additional security regions to be given to specific employees. The problem, however, was that this access was not temporary and in many instances was not removed, resulting in employees having access to far more data than their role would require. CoronIT didn’t even restrict data access to a certain location, nor was any separation in data access based on roles made.

It's also important to note that while medical professionals, as stated in Section 5.4.2, may argue that access to all data is required based on their role, based on the “old primary thinking of the infectious disease doctor, who must know exactly what the patient is getting and which treatment because he must be able to account for it afterward”, this is completely dependent on the specific role of the user and the expertise of the individual that will be working in any system. While this may apply to a specific role for a limited number of individuals, systems should not be designed based on this viewpoint.

Continuing on the previous requirement, the functionalities that users can access should also be based on their role. Higher-level functionalities should be restricted to a significantly more limited number of employees that require access to these functionalities, in the context of their role. For example, not all users should be able to make either export data from the database or be able to delete any records except possible dossiers they created themselves. This should only be available to specific roles that are for example better trained, and more closely monitored. The GGD case provides an example of why this is an important requirement, as stated In Section 5.3.2, as HPZone allowed each user to have full access to all functionalities, including an export function that was not monitored and could export significant amounts of data. Which has likely contributed to personal data breaches related to this system.

Continuing on the previous requirement, access to specific higher-level functionalities or operations should be restricted to require manual approval. If there is a significant risk to the data subject then even with access being limited to a limited number of individuals, it may not be a privacy-oriented design. However, if these functionalities require the approval of an even more limited number of individuals, or have a predefined waiting period where it is visible that such requests have been made, then this increases the security, while still being able to make use of these functionalities. One example of such a feature may be the exporting of data, either in a smaller capacity for internal use or in a significantly larger capacity when data is transferred to other organizations. This has been based on security standards of the vulnerability disclosure, as discussed in Section 6.1.4, where approval from the director is required to be able to access certain information or export it from the database.

Another requirement is to ensure that technical and organizational measures are implemented in the system to allow an organization to be able to comply with the rights any data subject has under the GDPR, which were previously discussed In Section 2.1.1. These rights are the right of access by the data subject (GDPR Article 15), the right to rectification (GDPR Article 16), the right to erasure (GDPR Article 17), the right to restrict processing (GDPR Article 18), the right to data portability (GDPR Article 20), and the right to object (GDPR Article 21).

Continuing on the previous requirement, a data subject can only be properly informed and data can only be correctly rectified if there is a clear overview of data use across the entire system. Which is currently lacking in many organizations, as indicated by Section 6.4. Although many of these should be included in the DPIA created for the system. As many systems are legacy systems that predate the GDPR and therefore would not be covered by it, this may be the reason that organizations don’t have an overview of data usage.

Another requirement is to ensure that data is not duplicated, being stored in multiple locations. Due to data being stored in multiple locations, errors will likely be created when data in one location is updated, without the data in other locations reflecting this change. When this happens, the data related to the data subject is inaccurate and needs to be rectified when it becomes apparent, according to the right to rectification (GDPR Article 16). However, the focus should be on preventing this in the first place.

Preventing errors related to the duplication of data was supposed to be addressed by the ‘basisregisters’, which are depicted in Appendix Section A5, Figure 60. The goal of this initiative was to function as the source of information for other organizations so that data can be adjusted in a central location, with changes then immediately taking effect in all organizations that use data from this source. Another advantage to this is that an individual is now able to communicate changes to a single location, instead of to multiple organizations, for example, based on an address change. If this information is not changed in every organization using this information, then personal data breaches may occur due to information not going to the intended recipient, which, as stated in Section 4.3.2, is one of the most occurring reasons for a personal data breach.

Another requirement is to properly back up data to ensure that no data is lost due to unexpected crashes, calamities, accidents, or any kind of malicious attack such as ransomware. Which possibly occurred with CoronIT, as discussed in Section 5.4.1. According to Z-CERT, an organization created by the Nederlandse Vereniging van Ziekenhuizen, the Nederlandse Federatie van Universitair Medische Centra, and the Nederlandse GGZ to investigate cybersecurity in healthcare, Ransomware is the biggest digital threat to the healthcare sector [198, p. 6]. This threat has been quickly growing in Europe, affecting 116 healthcare locations in Europe in 2021 [198, p. 7]. If unable to protect against a ransomware attack, a proper backup system is the only way to avoid either loss of data or paying ransom to malicious actors.

One potential problem in this is that the GDPR allows data subjects to use the right to erasure (Article 17), to erase their records. Opinions among supervisory organizations differ on this subject, although the consensus is that technical steps need to be taken to make it possible to delete individual records from backups or an organization needs to explain why this is not technically feasible to delete their records from these backups as well [199]. The AP's opinion on this matter is more clear, stating that data subjects do have the right to erasure on backups as well [200]. Backup systems need to be designed in such a manner as to be able to process requests for deletion and processing of them. There are exceptions to this however if the technology does not allow for the deletion of partial data, such as data stored on tapes, although on recovery of data, all data that has been requested to be deleted still needs to be deleted.

Another requirement is to make use of a common language, which was stated to be one of the main problems for the processing and sharing of data, as discussed in Section 6.4. While this requirement is primarily focused on the exchange of data between parties, it also has significant value in ensuring that employees will not make mistakes due to standards differing from what they are used to. To be able to provide efficient and safe healthcare, the information shared between all healthcare providers must have the same meaning or purpose, in addition to agreements describing the way technical terms are used in computer systems in the exchange and re-use of information [190, p. 3].

7.3.2 System Design

The monitoring of user activity is a requirement based on the NEN7513 norm, which requires the registration of either accessing or making adjustments to patient records [176]. However, for the recording of user activity to be effective as a technical measure to ensure the security of data processing, this norm should be extended to recording all system interactions. For example, the previous definition would not register search queries that would expose information that would also be valuable or risk negative effects to the data subject. It should also cover other processing related to this data, such as its use in other organizations for various purposes. It is also important that there are agreements specifying how and at what frequency this logging information is investigated, as well as who is responsible for this practice. Logging information should also be stored for a substantial time so that these investigations are possible In the first place. While this is a currently active legal requirement, the GGD case study shows, as described in section 5.3.3, that not only are interactions not logged in some instances, this data is never monitored either internally or externally as no protocols or agreements exist.

Continuing on the previous requirement, the logging of user activity is only valuable when this data is also being monitored. At a certain scale, the only way to be able to realistically monitor this is to do this automatically. This was stated as a solution to address the personal data breaches in CoronIT, as discussed in Sections 5.3.2 and 5.5.1. For example, flagging users based on suspicious behavior matching certain predefined conditions such as logging in outside of allocated time, logging in on a different device, accessing a significant number of files in a short period, opening patient records without any actions being taken, etc.

Another requirement is to make use of already existing practices and initiatives when creating the system, as it is unlikely that any newly developed system will be unique. Many governmental healthcare systems, or even governmental systems in general, have been based on decades of experience and would therefore contain elements that could be introduced in any other project. For instance, the government has created numerous data standards such as the BIO, as described in Section 6.1.1. Instead of creating the security standards from scratch, elements from the BIO can be introduced without being required to follow this norm. Section 6.1.2 offers another example of established practices that could improve security, which is to utilize an existing list of "people of interest" and flag any users where a search involved one of these individuals for manual review.

Another requirement is to protect the data from external access. While it is important to protect systems internally, security may still be compromised by any outside attack. Given the value of data contained in healthcare systems and the expected scale, any organization or system storing this data is a valuable target. Which requires organizations to be prepared and ensure that data can only be accessed via the system. While there was no indication of any external breach in the case study, and all organizations discussed in Section 6.1.4 placed a strong emphasis on protecting data from outside, these dangers should still be taken into account. As such, information should never be stored unencrypted, and organizational measures should be taken in every single location that has access to this data to ensure that no outside actors can gain access to any data.

Continuing on the previous requirement, protecting the data from external access will also require a secure working environment that users of the systems use to gain access to the data. If this is not done correctly then this becomes another avenue in which access to the system can be gained by malicious actors. As stated in Section 5.3.4, a secure working environment is prioritized over access being gained via accessing an URL, as using an URL would bypass any security measure that could be implemented in this environment. Allowing the use of non-company-issued devices also poses an increased security risk, as this prevents organizations from being able to ensure that devices are updated with the latest security updates, that the hardware itself is secure enough, and that there is no malware already present on the device that would compromise the system. However, this may not be a realistic requirement in a crisis and can be mitigated by taking other security measures.

Another requirement is for the system to be designed with the user in mind, in the context of UI/UX. While systems must be secure and useable in the context of being able to be used to support the process it was designed to be used for, this does not necessarily cover the way that systems are presented to the user. While it is acceptable for smaller systems to be presented more complexly as they will only be used by specialized individuals with intrinsic knowledge of the system, this should not be the approach used for the development of a new system. Instead, systems should be presented in such a manner as to be understandable with a minimal amount of training. While systems could be designed to support complex operations, there is no reason why the UI/UX itself should also be applied. An example of this can be found in the design of HPZone, as discussed in Sections 5.2.2 and 5.4.4, which was used to support source and contact tracing investigations. There should not be any complexity in the representation of which numbers need to be called as part of the investigation, or representing locations of contact, yet users indicated that it was a "fairly complex system, with several programs next to each other".

Another requirement is for the system to be interconnected with other governmental organizations, to a certain degree. Data may be used for research purposes in governmental organizations, as stated in Section 2.1.3, or it may be used by the RIVM to base its strategies on as previously stated. Figure 5 and Figure 6 in Section 5.1 provide another example of this interconnectivity, with systems and organizations being connected in support of a single purpose.

Another requirement is for the system to have enough built-in redundancy and capacity to function at all times. This is related to both the ability of the system to handle the expected demand from the number of employees processing data in the system and the amount of data that can be stored in the system. Although storage limits may also be addressed during the project, as long as this has been taken into account. While systems should ideally be accessible 100% of the time, this may be an unrealistic target. The governmental standard for the availability of governmental websites is 99.9%, or three-9 accessibility, as based on figures from 2008, although this level of accessibility was also not able to be reached [201]. While a governmental website is different from a service, this matches the industry standard sets by Google, Amazon, and Microsoft, which guarantees three-9 accessibility, or 99.9% uptime, with 8 hours, 45 minutes, and 57 seconds of downtime a year [202]. Therefore, the required uptime for a healthcare system will also be defined as at least three-9 accessibility or 99.9% uptime.

This is one of the most important requirements as loss of access has various significant consequences, endangering public health. The first consequence is that given a disruption of the system, individuals that rely on this system are aided by the service the system provided. This results in individuals either giving up or calling back at a later moment resulting in demand on the system increasing even further. As discussed in Section 5.4.1, this can result in individuals not only risking their health but public health as well as people may be unaware that they are sick. This may also occur if data is lost due to access being disrupted if no proper backups could be created.

Due to interdependency with other organizations and systems, disrupting access to one system may affect other organizations or systems as well. Using the GGD case study as an example, as discussed in Section 5.4.1, a disruption In CoronIT resulted in HPZone not being able to be utilized effectively as test results that would have been used to start source and contract tracing investigations were unavailable. As data in the system may also be used by organizations such as the RIVM to dictate health policy, the information must be correct and up-to-date, which would be hindered by systems becoming unavailable. Using the GGD case study as an example, as discussed in Section 5.4.1, system disruptions led to incorrect numbers being used by the RIVM to dictate the Covid-19 response strategy, as the number of positive tests was far higher than reported to the RIVM.

7.4 Specifications and requirements related to the processing stage

The processing stage refers to the stage where initial development has finished, although the development of certain elements or expansions to the current system may continue. The specifications and requirements related to this stage are mostly focused on the individuals that will be working on the system and the actual sharing of data from the system. The three areas that will be covered are the training of employees, other requirements for employees, and conditions and restrictions related to data sharing.

7.4.1 Training

Any system will require training materials that clearly describe the actions involved with any role, based on for example a use case diagram that has been combined with how these various tasks can be accomplished via the system. In addition to understanding how the system works, employees should also receive a basic understanding of the implications of processing medical information, as stated in Section 5.2.2 where certain actions are medical acts, and general training related the data security and GDPR, which are sometimes not given as stated in Section 5.4.4. This section also shows that employees would start working within a system while still finding the system to be complex and not being sure that they would be able to do their job correctly. Therefore, an additional requirement will be to require employees to demonstrate the effectiveness of the training through an examination. This will ensure that any employee working in the system is aware of both the activities related to their role, as well as the need and methods to keep data private and secure

Continuing on the previous requirement, the training process must be applied to every user of a specific role, independent of their location or conditions of employment. As users will be performing the same duties, according to the same protocols, in the same system, it is important to ensure standardization of the training. Without this, employees that are contracted from external organizations likely will receive a lower standard of training, affecting data security. As stated In Section 5.4.4, where contractors received different training materials, given that the employees giving the training themselves would have limited experience, leading to actions counter to the GDPR as they weren't even aware they were breaking it in the first place.

Even if the previous two requirements related to training are successfully implemented, it is unlikely that employees will not have questions related to their role. Therefore, organizations should anticipate this and provide an environment where employees can ask and answer questions in a secure and monitored environment.

The need for such an environment, in general, can be seen both at the GGD and U-diagnostics, as discussed in Section 5.5, where WhatsApp groups were created with large groups of employees. WhatsApp however, is an insecure platform that is unfit for handling medical information given that even if deleted, information may already be stored on the device of everyone in the WhatsApp group. To make matters worse, questions were asked that contained uncensored medical information, which constitutes a personal data breach. Disallowing this practice without providing a secure environment will not work, as discussed in Section 5.6.5, as these WhatsApp groups fulfill a clear need and would only result in such groups being recreated where organizations are unable to monitor them at all.

7.4.2 Employees

The first requirement related to employees is connected to the requirement for a secure access environment, as stated in Section 7.3.2. While such an environment may be secure from outside attacks, any actor could still gain access via the login credentials of employees. Which is something that can not solely be addressed from a system perspective. Two problems that require user involvement that must be addressed for this requirement are the conditions around 2FA use and lockout times.

Two-factor authentication, or 2FA, refers to an authentication method where two individual and separate methods are used to gain access to a system. For example, the user first enters their user credentials and then approves the log in via a second method such as approval via their phone. While this is a purely technical solution, it can be compromised without any technical method. If a user's credentials are ever exposed, a malicious actor can use these credentials to pass the first method. This would then be blocked by the requirement to approve this attempt via a second method, however, if the user then approves the login attempt by mistake or a malicious actor has gained access via the device, the system is still compromised. Therefore, employees need to be informed about any such security method so this scenario can be avoided.

The second way to gain access to the system does not require any credentials, instead, a malicious actor or curious actor can make use of a terminal that has not yet broken its connection with the system. While lockout times can reduce the chance of this scenario occurring, they can't prevent it entirely as the lockout time needs to be large enough to not interfere with normal operations. If such systems may be able to be accessed from home, the threat of this is significantly higher. Therefore, users need to be informed that they need to log out of this system to prevent this scenario. Even when working from home, the user should never leave their computer behind unlocked in addition to other measures to ensure that the home workplace is as secure as the office, as stated by the German Federal Office for Information Security [203].

Another requirement for employees is to require the use of instruments such as a VOG, to filter out malicious actors from gaining access to the system, and to require the signing of a non-disclosure agreement, to reduce the likelihood of employees causing personal data breaches and keeping the data contained in the system private and secure.

Specifically, a VOG should be required before an employee is allowed to gain access to the system as Section 5.4.3 shows that using any other method using the VOG results in an approach that is not secure. If an employee can gain access based on the condition to later submit their VOG, people that would not be able to gain a VOG would be able to gain access anyway. If using a contract where the user signs that they would be able to get a VOG before being able to gain access to a system, an organization implies that this document would have the same value as a VOG, which it clearly wouldn't as one is based on an individual's declaration and the actual VOG is based on a governmental declaration. It is also of critical importance that this requirement is validated for every employee, as otherwise, people that do not have either a valid VOG or never submitted one in the first place would still be able to gain access to the system.

Requiring a non-disclosure agreement may not ensure security, as proven by the significant number of personal data breaches in the GGD case study where all employees had signed such a document, but it is an easily implementable and affordable measure that may be able to improve security.

Another requirement is for the system to inform the employee of (part of) the implemented security measures and sanctions related to breaching the privacy and security of data in the system. This has two major benefits, of which the first is deterrence and the second is to prevent the employee from defending his actions by claiming ignorance. Ideally, these messages should be displayed to the user every time the system is accessed.

As employees are made aware of (part of) the implemented security measures, they will be less likely to attempt any malicious activity as they are now aware that they are unlikely to be able to circumvent them, acting as a deterrent. It is important to note that while this informs the employee of security measures, there is little to no risk to this, unlike the GGD statement made in Section 5.3.3 where security measures are not described due to the effectiveness of these checks. System security should not depend on the secrecy of the implementation or its components, and if it does, then it is likely that the security measures were never able to increase security at all. The inclusion of sanctions would make it less likely that employees would try to circumvent the security measures, even if they believed they were able to.

As employees are made aware of the sanctions related to breaching security, they will also be unable to claim ignorance of their mistakes. Making it more likely that any action that would result in a personal data breach is not due to a mistake, but instead a malicious activity that can be treated accordingly.

Continuing on the previous requirement, the previous requirement could also be combined with the requirement to confirm certain actions to prevent mistakes and to be able to display an informational text informing the user of certain security measures and sanctions if the action is taken without any valid reasoning. This minimizes the probability of employees accessing data or using functionalities by mistake, while still allowing the employee to perform their role if the action does have valid reasoning. The combination of the informational text aims to reduce the number of mistakes made by employees as they will be more careful in their actions and it may deter malicious actors as they are aware of the consequences when their actions have no valid reasoning. This is a practice already used in the healthcare sector, as stated in Section 6.1.2, where healthcare employees need to confirm access to patient files.

7.4.3 Data Sharing

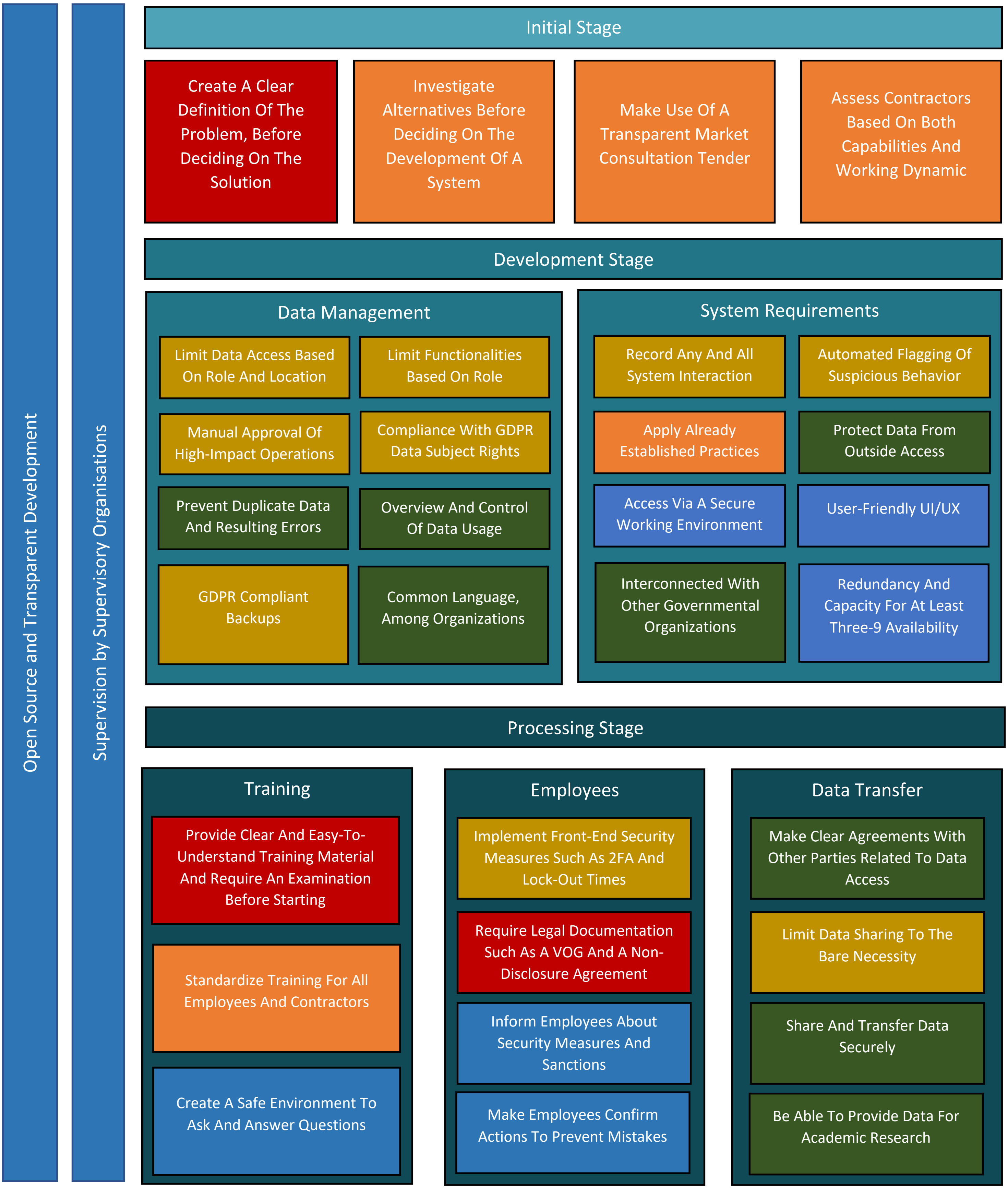
When data needs to be shared with any organization outside the confines of the current system, clear agreements must be made with the other parties related to how the data is going to be used, and protected, how personal data breaches will be addressed, and how data will be disposed of after its use. Given that data that leaves the system results in the original organization losing all control over it, these agreements must be not only made but also monitored. Attention must also be paid to which party would receive the data, to ensure that technical and organizational measures are in place at the receiving party to be able to adhere to the agreement.

Continuing on the previous requirement, the data that is shared with any other organizations needs to be shared securely, without being able to be intercepted or possibly resulting in a personal data breach in any other way. Data should be encrypted and only be able to be accessed by receiving party. Depending on the way that data is shared, either digitally or physically via the transport of a physical copy, additional measures need to be taken to ensure the security of this transport.

Continuing on the previous requirement, data must be shared in a manner that results in the least impact on the data subject. The decision must also be made if either the data itself should be shared or if statistical information about the data, in the form of metadata, could be sent instead. If the actual data is sent, then this should either be done in such as way as to make the information impossible to trace back to the data subject, or if this is not possible, to at least minimize the impact such data can have by restricting certain columns.

Even if the previous three requirements related to data sharing are successfully implemented, additional organizational and technical measures need to be implemented to allow for data to be shared in the context of academic research too. This appears to be a problem given that as discussed in Sections 2.1.3 and 5.6.9, organizations are hesitant to make their information available to academic parties, not for any technical reason, but for legal reasons where organizations are afraid of violating the GDPR by sharing data.

7.5 Overlap with the Proposed FAIR-Based Framework



Open Source and Transparent Development

Supervision by Supervisory Organisations

Initial Stage

- Create A Clear Definition Of The Problem, Before Deciding On The Solution
- Investigate Alternatives Before Deciding On The Development Of A System
- Make Use Of A Transparent Market Consultation Tender
- Assess Contractors Based On Both Capabilities And Working Dynamic

Development Stage

Data Management

- Limit Data Access Based On Role And Location
- Limit Functionalities Based On Role
- Manual Approval Of High-Impact Operations
- Compliance With GDPR Data Subject Rights
- Prevent Duplicate Data And Resulting Errors
- Overview And Control Of Data Usage
- GDPR Compliant Backups
- Common Language, Among Organizations

System Requirements

- Record Any And All System Interaction
- Automated Flagging Of Suspicious Behavior
- Apply Already Established Practices
- Protect Data From Outside Access
- Access Via A Secure Working Environment
- User-Friendly UI/UX
- Interconnected With Other Governmental Organizations
- Redundancy And Capacity For At Least Three-9 Availability

Processing Stage

Training

- Provide Clear And Easy-To-Understand Training Material And Require An Examination Before Starting
- Standardize Training For All Employees And Contractors
- Create A Safe Environment To Ask And Answer Questions

Employees

- Implement Front-End Security Measures Such As 2FA And Lock-Out Times
- Require Legal Documentation Such As A VOG And A Non-Disclosure Agreement
- Inform Employees About Security Measures And Sanctions
- Make Employees Confirm Actions To Prevent Mistakes

Data Transfer

- Make Clear Agreements With Other Parties Related To Data Access
- Limit Data Sharing To The Bare Necessity
- Share And Transfer Data Securely
- Be Able To Provide Data For Academic Research

Figure 10 - High-level Overview of Specifications and Requirements in Relation to the FAIR-Based Framework

- Part Of FAIR By Default
- Part Of FAIR Based On Its Implementation
- Extension to Systems Built Using the FAIR-Based Framework
- Adressed Through The FAIR-Based Framework
- Not Included In The FAIR-Based Framework

7.5.1 Overall Project

The aspects of open source and transparent development and supervision by supervisory organizations, as discussed in Section 7.1, are fully compatible with the proposed FAIR-based framework. These aspects are intended to ensure that technical and organizational measures are implemented correctly, which increased compliance with the GDPR and increased the likelihood that people will perceive any such system as secure. Meeting the specifications and requirements as discussed in this section depends on its implementation in any system, with these aspects acting as a way to both monitor that this has been done correctly and to be able to intervene when this has not been done correctly. Additionally, it should be avoided that any system would be regarded as safe, just because it has been constructed using the FAIR principles. As this alone is not enough to ensure that any such system would actually meet the specifications and requirements, and legal requirements, in the healthcare sector or beyond.

7.5.1.1 Open Source and Transparent Development

The current application of FAIR, being an open and scientific standard, already meets many of the ideas behind the concept of open-source development, increasing the amount of transparency in the processing of data. However, this application has currently not been expanded beyond the processing of data, for example expanding its usage towards the description of organizations and relations between those organizations or its possible value in the creation of systems themselves. The proposed FAIR-based framework aims to address this through a structured governmental ontology containing this information.

The application of FAIR is not based on creating new standards, but instead is based on using and expanding existing ones, which is similar to the more general open source concept where a community build upon previous work in an ever-improving product or service. The proposed FAIR-based framework aims to continue this approach, continuously expending the ontology with additional information that can be used in the service of processing data or gaining more control and transparency in organizations, projects, and the relations between each of these entities. The FAIR standard itself, being created as recently as 2016, also remains in active development, based on the contributions of the FAIR community.

Opening up (parts of) both the development and use of projects and the wider governmental ontology has the beneficial effect of being able to monitor this development, and providing valuable input, thereby increasing the probability of success. The additional exposure of information that is currently not accessible to the public, in a structured and easily accessible way, is also likely to lead to better decision-making as any decision is now open to public scrutiny. As proven by the impact of public appearance in the context of political involvement in the case study and the effectiveness of the AcICT even though following their advice is not mandatory, this form of transparency has a significant effect on decision-making. Additionally, this approach allows for example FAIR experts to contribute to the development process, which may prove vital to its success given the lack of experience with FAIR in both the government as well as any contractors that would be hired

7.5.1.2 Supervision by Supervisory Organizations

While the proposed FAIR-based framework is unable to address any of the funding and/or capacity-related obstacles present in the supervisory organizations, the way in which supervisory organizations are used, and thus their effectiveness and involvement in project development, could be changed based on the implementation of the FAIR-based framework. This is achieved through the implementation of the governmental ontology, covering not information related to the processing of data within systems, but also covering information other relevant information related to projects and organizations. Through this information, the usage of supervisory organizations, as specified in Section 7.1.2, can be focused on the most important projects instead of using a sampling-based method in the context of the AcICT, and the AP can get involved in projects before reports are received based on a failure to meet certain criteria.

In addition to being able to more effectively select projects for further evaluation, the AcICT is also likely to monitor the creation of the default version of the FAIR-based framework used for system deployment. This involvement is due to the fact that FAIR, being a new technology that neither the government nor contractors would have any experience with, would warrant such an evaluation to increase the probability of success. As this default version is used to create projects in the future, such an evaluation would be able to improve governmental ICT systems far beyond this initial evaluation. Additionally, monitoring a FAIR-based project, in general, may lead to lessons being able to be applied in other projects, also improving governmental ICT systems in general. Through this, the second primary task of the AcICT is covered.

The AP is also likely to be required by law to be involved with the creation of this default framework, as well as the first project(s) developed based on this default framework. As this would be the first application of the FAIR-based architecture, with FAIR being a new technology which no data protection impact assessment has been carried out for. Under Recital 89 of the GDPR, any organization using new technology will be required to notify the supervisory organization and submit this DPIA for assessment [204]. Which requires both the creation of a DPIA and leads to the involvement of the AP in this process.

7.5.2 Initial Stage

Of the four specifications and requirements from the initial stage of a project, the proposed FAIR-based framework aims to address three of these aspects. These are primarily addressed through the increase in information availability related to organizations and the use of projects and systems across the dutch government. The only aspect unable to be addressed through this proposed FAIR-based framework is the initial investigation to decide upon the problem itself and how this should be addressed. Without meeting all these aspects, it is unlikely that any project will succeed, but for this proposed framework, it will be assumed to this investigation has been conducted properly and that the solution to the problem requires the development of an ICT support system.

In the context of investigating alternative projects meeting the specifications and requirements of a project, creating an overview of the usage of projects in systems in other organizations would make it significantly easier to first of all identify any such alternatives based on the purpose of these systems and the data that they process. When such alternatives have been identified, the increased amount of information related to these projects and systems would also make it easier to, at least initially, evaluate if these potential alternatives meet the security and regulatory requirements. Any such alternative that meets these requirements can then be investigated in even more detail, beyond the information contained in the ontology.

In the context of making use of a transparent consultation tender, the proposed FAIR-based framework is unable to ensure that this process is followed, but the information included in this ontology can both monitor if this process has been followed and if not, based on which legal ground such a process has not been conducted. In the context of the covid pandemic, this was possible due to an exception in the aanbestedingswet as stated in Section 5.2.1, which could reference this specific exception. However, if no such reference has been created, and no such exception has been used, this could lead to this decision being legally challenged. Even if such an exception has been used, it could also be legally challenged if the use of this exception is legally valid.

In the context of assessing contractors based on both capabilities and working dynamics, the information included in the ontology could be used to identify trends based on previous performance. Which could identify projects that are possible beyond the capabilities of any specific organization, or it could be used to identify which organizations would be most suitable to develop a project based on various information. It should be noted that this approach would not function for organizations that haven't previously conducted any, or a low number, of projects for the dutch government, although it could also be argued that such an organization would not be the most suitable either in a crisis situation or beyond, or would need to be assigned smaller projects to prove that they have both the capability and working dynamic to successfully complete a project for the Dutch government.

7.5.3 Development Stage

All of the specifications and requirements related to the development stage can be met by a FAIR-based architecture. However, while FAIR addresses all technical issues, it does nothing to address organizational decisions. While a FAIR-based architecture can be created in such a way as to provide a secure system that adheres to data minimalization and privacy by design, it is also possible to create a FAIR-based architecture that would meet none of these requirements while still being FAIR-based. This section will cover the specific specifications and requirements in more detail, including the considerations that must be made to make the FAIR-based architecture compliant with the GDPR.

7.5.3.1 Data Management

While FAIR can control and limit data access by role and location on a technical level, this is dependent on its configuration. Any FAIR-based architecture will face the same organizational problems, that the decision to limit data access is either made incorrectly or not at all. Due to the customizability of FAIR, you can create a system where everyone has complete access to everything or no one has any access. The first one is unlikely to adhere to the requirements of privacy by design and by default, the second is unusable. Therefore, FAIR can comply with this requirement, but like the original situation, it depends entirely on its implementation.

The same argument also applies to the requirement to limit access to functionalities based on roles. A FAIR-based framework can implement these limits on a technical level, but it relies on the organization to implement them. However, a FAIR-based approach may be able to make some functions obsolete, which would improve security. For example, any function related to the exporting of data can be removed as data is not exported in FAIR, instead, data access is granted.

The same argument also applies to the requirement to manually approve high-impact operations. A FAIR-based framework can implement manual approvement of data access or functionalities on a technical level, it relies on the organization to determine which operations are high-impact operations. However, a FAIR-based approach may be able to make some high-impact operations obsolete as one of the examples that were provided on a high-impact operation was the exporting of data which can be removed as data is not exported in FAIR, instead, data access is granted.

Compliance with GDPR data subjects rights does not differ from a non-FAIR-based framework, requiring technical measures to make the system able to act in accordance with these rights and organizational measures to be able to process requests and make use of these technical measures.

However, a FAIR-based framework may be able to improve upon this aspect due to a heightened level of data control. As the organization is always in full control of its data, due to never sharing any files beyond the control of the organization, a FAIR-based approach would be able to ensure that all data related to a data subject can be appropriately returned, restricted, deleted, or transferred.

In addition to this, giving data subjects access to the FAIR network directly allows for a superior approach compared to a non-FAIR-based one. Through this network, a data subject could be securely given access to their data and exercise their additional rights, without needing to first request this from any organization, bypassing traditional, slower, methods. It would also allow for a data subject to provide access to their data for various academic purposes. As while the GDPR may provide certain exceptions where the data subject does not have to be consulted about the use of their data, at least allowing the data subjects' opinions to be considered would be an improvement and eliminate organizations' hesitancy in providing access to data based on the reasoning that the data subject was not consulted.

Preventing duplicate data and the resulting errors is a property of a FAIR-based framework by default as data is meant to be stored at a single location, with organizations being able to provide access to that data. Through this approach, updates to data would only be processed in a single location, which eliminates the possibility of differing copies. However, it is important to take into consideration that updates to data would not be applied to any backup of the data. If data ever needs to be restored, then the information in this backup would be incorrect unless a separate backup is made that stores the various transformations on data which could then be applied again.

The requirement to have a clear overview of data usage and control over it in the organization is a property of a FAIR-based framework by default as stated in Sections 2.2.3 and 2.2.4. Due to the data visiting approach to giving access to data, an organization has a comprehensive overview of who has been provided with access to data., with the assurance that data hasn’t been shared outside of the confines of the system where it would be beyond the organization's control. Extensive metadata connected to the data would be able to clearly describe what data is, what it is used for, who collected the data, what the policy surrounding access is, and what duration the data will be stored for.

The requirement to properly back up data to ensure that no data is lost due to unexpected crashes, calamities, accidents, or any kind of malicious attack such as ransomware rights does not differ from a non-FAIR-based framework, although a FAIR-based framework may be more vulnerable unless measures are taken to address this fact. As there is no other location with a copy of specific data, it is of critical importance that the backups at that location do not become compromised as that would lead to an unavoidable loss of data. While a traditional approach faces this same problem, there is at least the chance that another location may have a backup of at least part of the data. Deleting data based upon the request of a data subject will also require technical and organizational measures to be taken, but these measures don’t differ from a traditional approach.

The requirement for a Common Language, among organizations, is a property of a FAIR-based framework by default as stated in Section 2.2.4. The exact language is not important, as long as it is a common language shared between all organizations. This can be achieved either through the creation of a new standard which could be created for the Netherlands or the entire European Union, or the use of an existing standard. The second method may be more valuable as it would immediately make all data processed using the FAIR-based framework usable for anyone already using that standard if access is provided at least. The most important aspect of this is that a standard is chosen that as many organizations as possible will use, which could be extended to use healthcare data not only in health facilities or research facilities in the Netherlands but would allow the entire world to use data from the entire world without any difficulty.

7.5.3.2 System requirements

The requirement to record any and all system interaction is not a property of a FAIR-based framework by default, depending on its specific implementation to meet this requirement. However, as a FAIR-based framework provides organizations with complete control over their data through the data-visiting approach, implementing this requirement is a relatively minor addition. By recording the activity of either organizations that have been given access, or users that have been given access including their specific role, a comprehensive monitoring system is created. Connecting this component to the secure working environment would enable even more monitoring, beyond the limitations of a web browser. The final requirement would be to create agreements to investigate these recordings, as just recording data does not necessarily result in a safer system.

The requirement to automatically flag suspicious behavior is not a property of a FAIR-based framework by default, depending on its specific implementation to meet this requirement. However, given that the previous requirement is met, this is a relatively simple addition that would only involve the creation of indicators and queries to be able to flag any user that shows what the organization would define as suspicious behavior. Correct implementation of this requirement is realistically the only way to be able to properly monitor systems at this scale, significantly reducing the required amount of resources for manual checks while not decreasing or even increasing the likelihood that suspicious behavior is detected and the correct users can be investigated. The effectiveness of this requirement relies entirely on its implementation, however, meaning that just because suspicious behavior is flagged, doesn't mean that malicious users are always able to be detected or that they could be detected quickly before being able to create a more major personal data breach.

The requirement to apply already established practices is not part of a FAIR-based framework by default, nor is it dependent on any of the FAIR concepts. It is however an important requirement as it is unlikely that any system used to support a healthcare process will be truly unique, nor are systems developed in a vacuum. Due to this, established practices have already been created and implemented in other systems to address various challenges and problems. The proposed FAIR-based framework aims to promote this application by including these practices in the common default base used to develop a project, and through monitoring the inclusion of these practices on a project-specific level in a comprehensive ontology. One example of this has already been discussed in Sections 6.1.2 and 7.5.3.2, which is to make use of a list of individuals of interest that would automatically flag any individual that accesses their patient dossier, for use in the previously stated monitoring system. But various technical requirements, as stated by the BIO or any other standard, could also be used to both monitor and incorporate into the common default base.

The requirement to protect data from outside access is a property of a FAIR-based framework by default, as without meeting this requirement many of the advantages of FAIR would be negated. It does not mean that this does not need to be carefully implemented, however, as this requirement will require careful implementation and continuous updating as new vulnerabilities need to be addressed. Adhering to this requirement requires both organizational and technical measures as while systems themselves may be secure, any individual working in an organization could commit actions that would result in that organization being compromised, bypassing various implemented security measures. Preventing this would require training related to security, for example, training aimed to be aware of and prevent social engineering, and/or training aimed at using secure passwords and not connecting unknown devices to anything connected to the organization network.

If this requirement is not met then the entire concept of controlled access via a FAIR-based framework would be able to be circumvented, which would mean that an organization would lose control over its data and be at risk of both major personal data breaches and disruptions to the operations of the system, based on the types of attacks that occurred due to a failure to secure the system and the data within it.

The requirement for access to be provided only via a secure environment is not a property of a FAIR-based framework by default. It could be considered a property of a FAIR-based framework on its implementation, based on the property of “Accessibility”, however, this thesis has chosen to consider this as an extension instead. While any FAIR system can make use of a working environment in the form of a client, or a web interface, these environments aren’t necessarily secure. To address this, it is the platform itself that will need to be integrated with security-related improvements, instead of any of the underlying infrastructure or technology. The implementation of which is based on utilizing a client instead of a web interface, enforcing the usage of 2FA and stronger passwords to access the client, requiring any connection to originate from the secure working environment client, and using additional information gathered by this client to monitor employees and intervene when necessary.

The requirement for a user-friendly UI/UX is not a property of a FAIR-based framework by default, nor can it be considered a property of a FAIR-based framework on its implementation. Instead, this requirement should be considered as an extension, focusing on improving the platform itself, instead of any of the underlying infrastructure or technology. Given that FAIR is a scientific standard, used primarily by data scientists, the focus would have been on being able to perform all required operations, which are often complex, in favor of focusing on usability. However, these concepts also don’t exclude each other, with complex operations being able to be performed using a simple UI. This requirement is especially important for supporting the primary care process, as new employees are unlikely to be well trained or experienced with using this or any comparable platform.

The requirement for the system to be interconnected with governmental organizations is a property of a FAIR-based framework by default, based on the data visiting approach where access to either data or some derivative of that data can easily be provided to any organization. However, it is important to carefully evaluate which form of access should be given to governmental organizations to support various processes. In most instances, providing statistical information about the data would be sufficient for their purposes. For example, an organization such as the RIVM would primarily have to be provided the daily figures or trends, which they can then use to support the central government by creating advice based on this information. Other organizations such as the CBS, a trusted third-party organization, would most likely require a form of access where they could request to run their algorithms in support of research.

The requirement to ensure redundancy and capacity for at least three-9 availability is not a property of a FAIR-based framework by default, nor can it be considered a property of a FAIR-based framework on its implementation. Instead, meeting this requirement is considered an extension and is based on investments into the required amount of computational resources to handle the expected amount of traffic, the required amount of storage capacity, and redundancy to be able to address any temporary outages in any specific system.

The federated approach to data management, divided over multiple locations, functions as some form of redundancy and availability for the entire network, but would not be able to create redundancy and availability for that specific location. If the connection with this location is disrupted, then all other locations would be able to function normally, unless something in the core of the network would be disrupted. As data is not duplicated and stored only at that single location, no other location would be able to temporarily take over the tasks of the disrupted location.

To ensure redundancy and capacity for at least three-9 availability, every location will have to individually meet these requirements. This will require any system to take into account the expected amount of traffic when designing the system, and acquiring the required amount of computational resources to handle this traffic. It will also require any system to take into account the maximum storage capacity, with this capacity either being acquired or the system being constructed in such a manner as to be able to scale this capacity further when the need for storage increases. To prevent disruptions to the workings of any facility, multiple redundancies are required for any specific component. Finally, an extensive backup solution is required to ensure that systems are not vulnerable to data loss due to mistakes, hardware failures, or ransomware attacks.

7.5.4 Processing Stage

Most of the specifications and requirements related to the processing stage can be met by the FAIR-based Framework. However, in most of these aspects, a purely technical solution does not fully address the problem, requiring a significant amount of organizational involvement to ensure these specifications and requirements are fully adhered to. This applies especially to the training of employees and employees in general, where technical solutions are used, but their effectiveness is determined by how an organization has chosen to implement them. The transfer of data is a primarily technical issue that is completely addressed by FAIR, however, even in this case organizational decisions may still result in too much information being shared or information not being shared at all. This section will cover the specific specifications and requirements in more detail, including the considerations that must be made to make the FAIR-based architecture compliant with the GDPR.

7.5.4.1 Training

The requirement to provide clear and easy-to-understand training material and require an examination before starting is not part of a FAIR-based framework by default, nor is it dependent on its implementation. It also cannot be considered an extension to the FAIR-based framework or be addressed through an increase in information accessibility. Even the usage of a common default base, which would then require project-specific implementations that reduce the value of default training materials, or the usage of previously created projects that have related training material, is unable to ensure that the training material itself is clear and easy to understand. Neither can this approach ensure that employees undergo an examination before starting the processing of data in any system. Therefore, this requirement has not been included in the proposed FAIR-based framework. However, it remains a critical requirement and is closely connected to the requirement to have a User-friendly UI/UX, as the level of complexity of the interface has a direct impact on how difficult it will be to train employees.

The requirement to standardize training for all employees and contractors is not part of a FAIR-based framework by default, nor is it dependent on its implementation. However, it can be positively impacted by the common language aspect and the data federation aspect of FAIR. Given this federated approach, with projects being able to be deployed at multiple facilities and sharing the same code bases while having their own employees and own database, employees at any of these facilities would be able to receive the same training and instructions as these are standardized by default. However, while training material could be shared easily, it is up to the organizations themselves to ensure that organizational measures are taken to ensure that employees have been provided with the time to train, in addition to possible structured classes going into further detail. Given the fact that already existing projects can be used, training material would most likely reflect how the healthcare sector already operates, which would result in training materials already existing and the operating standards matching current experience. Although this may also be the cause of major issues as described in Section 5.4.2, where previous experience at the GGD led to some unwise decisions.

The requirement to create a safe environment to ask and answer questions is not part of a FAIR-based framework by default, nor is it dependent on any of the FAIR concepts. It does however connect to the data visiting approach as without such as environment, data within the system would be likely to leave its confines. When this occurs, an organization would lose control over it which would not only be able to lead to personal data breaches but would also prevent an organization from fully complying with all data ownership-related rights of the data subject as an organization no longer has an overview of who has access to that data and it also can’t be effectively restricted or deleted.

7.5.4.2 Employees

The requirement to Implement front-end Security measures such as 2FA and lock-out times is not a property of a FAIR-based framework by default, depending on its specific implementation to meet this requirement. However, implementing this requirement is important to ensure that access can be limited to the agreements that have been made between the FAIR-based architecture and employees or organizations that have been granted access. This access has been provided based on the conditions that the user or organization is who they say this is, which will require front-end security measures and lock-out times to minimize the possibility that any entity can get access that they would otherwise not be provided with. Implementing this is a relatively simplistic addition to the dashboard/platform.

The requirement to require legal documentation such as a VOG and a non-disclosure agreement is not part of a FAIR-based framework by default, nor is it dependent on its implementation. It also cannot be considered an extension to the FAIR-based framework or be addressed through an increase in information accessibility. Making this a standard requirement of the common default base could increase the usage of these documents, but it would ultimately be up to the organization to ensure that they both require these documents and follow this procedure correctly to ensure the validity of these documents and the overall process. Additionally, its requirement in any project could be included in an informational overview, but this would suffer from the same problem where you cannot be sure that the process around these documents is conducted correctly. The GGD case offers a practical example of this, with VOGs not being required until employees have already been provided access to systems for six full weeks, with some employees never being asked at all.

It should be noted that this remains an important requirement, as inside malicious actors remain a vulnerability of the FAIR-based framework. While information can be limited, it can never be limited to such an extent as to be completely safe as that would most likely result in a system that is not useful to support a healthcare process. Any monitoring method would be able to decrease the scope of a possible personal data breach, but it would be unable to prevent any from occurring.

The requirement to Inform employees about security measures and sanctions is not part of a FAIR-based framework by default, nor is it dependent on any of the FAIR concepts. While this is technically an organizational measure, a possible technical implementation would be to make this part of the login procedure as discussed in Section 7.4.2. Implementing this is a relatively simplistic addition to the dashboard/platform.

The requirement to make employees confirm actions to prevent mistakes is not part of a FAIR-based framework by default, nor is it dependent on any of the FAIR concepts. The exact implementation of this requirement does not differ from its implementation as described in Section 7.4.2 and would be a relatively simplistic addition to the dashboard/platform.

7.5.4.3 Data Transferring

The requirement to be able to make clear agreements with other parties related to data access is a property of a FAIR-based framework by default. Through the data visiting approach and included metadata specifying the conditions for access, clear agreements can be made with any organization. Agreements that in most instances cannot be violated as an organization would not be able to gain or maintain access beyond the conditions of the access agreement. It should be noted however that for this to be fully effective in ensuring security, agreements should be made with regard to the sensitivity of the data.

The requirement to limit data sharing to the bare necessity is not a property of a FAIR-based framework by default, depending on its specific implementation to meet this requirement. While FAIR allows significant control over the data being made accessible to other organizations, as well as the ability to run algorithms on data that hasn't even been shown to the user, it is also possible to provide access to the entire dataset. Therefore, organizations need to carefully consider which data may be made accessible and to whom.

The requirement to be able to share and transfer data securely is a property of a FAIR-based framework by default. In most instances, data will never have to be transferred using the data visiting approach, which is an inherently more secure approach to data management. The only instance in which data would be required to be transferred is when data from multiple sources need to be combined as discussed in Section 11.3.1.3, which could be done securely by transferring it to another location from which the data visiting approach could then be applied to provide access to an organization that would require this type of access.

The requirement to be able to provide data for academic research is a property of a FAIR-based framework by default. As access to the actual data does not have to be provided to an organization for them to be able to conduct an academic study, this should address the organizational and technical measures required to be able to allow for data to be shared in the context of an academic study. This should address the hesitance of organizations to make their data available as this could be done without sharing any data, which would most likely not constitute a violation of the GDPR even if Article 89 could not be utilized. The addition of a technical measure for data subjects to be able to easily provide or restrict their data for any purpose would further increase an organization's willingness to provide access to their data.

8. Proposed FAIR-based Framework

This section presents an overview of the proposed FAIR-based framework, based on the specifications and requirements of a GDPR-compliant healthcare system as discussed in Section 7. The overall aim of this Framework is to utilize FAIR in such a way as to not only be able to improve the sharing and use of data on a dataset level, but to extend this in such a way as to improve the way that entire projects and systems are constructed and utilized. This is achieved through introducing various aspects of the GDPR in the form of an ontology, covering facility/organization-specific information, as well as project/system-specific information, as well as GDPR-level data variable-specific information. Additional concepts that are used are data federation, data visitation, and machine accountability, which together with the previous concepts facilitate the development of systems adhering to data minimization and privacy by design and default principles.

For the proposed FAIR-based framework to be effective, it needs to be designed in such a manner as to be highly customizable, allowing its application in various crisis scenarios and data types. Considering the unpredictable nature of crises, any type of more rigid framework would either not be able to be utilized well in a possible future crisis, or it would require the creation of numerous different frameworks accounting for numerous different possible future crises. The probability of this alternative method succeeding is low as this approach would require significant investments to be made, preparing for situations that may or may not occur in the future, without the certainty that any one of these frameworks would be effective in the next crisis.

As the proposed FAIR-based framework is intended to support system development for any crisis, with each crisis being distinct and likely to necessitate the use of a different system, the primary goal is to develop a common standard used to support project-specific development. This is achieved through a common standard incorporating pre-developed main components, while at the same time allowing for project-specific adjustments to be created, capitalizing on the framework's generalizability and customizability. This is then further supported by the intention for different projects, that have already been completed, to act as an alternative starting point for the development of any new project. This has numerous benefits such as reducing the required development time, being able to incorporate implementations and improvements from other projects into new projects, and creating a large number of individuals that are proficient in working with the proposed FAIR-based Framework.

The proposed FAIR-based framework also aims to not only comply with the GDPR, but to improve upon traditional approaches to data access and allow for an improvement in the way that data subjects can exercise their GDPR-related data ownership rights. In the context of the approaches to data access, the most significant change is the fact that data access should revolve around the required level of information required to fulfill a task, instead of data access being provided to complete data values by default. To increase the privacy and security of data even further, organizations need to be in full control over their data, to be able to ensure automatic monitoring over all instances where data from a system is accessed. Additionally, when queries need to be run on a large number of individuals, this framework proposes this be done through aggregational statistics, to reduce or eliminate the need for any data related to any specific individual to be exposed.

Expanding the proposed FAIR-based framework to cover scientific data analysis was investigated extensively, however, it was ultimately decided to leave this out of the scope of this thesis. Instead focusing entirely on the processing of data in the primary process of supporting healthcare activities, as well as aggregational statistics. This decision has two main reasons, with the first being that the inclusion of supporting scientific data analysis is not essential to addressing a crisis situation beyond the more limited information that aggregational statistics can provide. The second reason is technical in nature, as while the methods and techniques specified in this section would offer substantial value to supporting scientific analysis, the additional requirements to do this properly, and numerous technical challenges would need to be addressed.

However, given the significant value of scientific data analysis, future work may explore this aspect in conjunction with the common standard. Section 11.2 outlines potential future work and highlights the primary issues of data storage and significantly increased computational requirements. To support scientific data analysis, data from multiple data sources would likely need to be combined, based on a certain key. This would require some form of data transfer to a specific location, which this framework has intended to prevent, due to the inherent risks for privacy and security to the data subject as well as the added level of complexity this would have on the monitoring of data access. The significantly increased computational requirements, based on the need to run computationally heavy algorithms and store derivates of forms of data, are also challenging as the decision to limit data storage to the facilities themselves would make it difficult for each facility to have the experience and resources required to be able to support this process.

Section 8.1 describes the content and design of a large-scale ontology, allowing for the future creation of what this thesis has called the Dutch governmental ontology. This, to be created, ontology is used to standardize and record relevant information and relations on three distinct layers, describing the information and relations related to organizations, the projects used and/or developed by these organizations, and the data variables used by these projects. The organizational layer makes use of and expands a previously created RDF triple based ontology created by the Dutch government. The project layer requires the creation of a new ontology, although some form of standardization has already been applied to capturing this information. The project variable layer makes use of an existing ontology to capture medical information, using SNOMED CT NL, which this thesis extends with GDPR-related information related to the processing of information by employees. While the focus of this thesis has been on healthcare-related systems, there is no practical obstacle to expanding this approach to any system, with this information being able to be included in the Dutch governmental ontology.

Section 8.2 describes several potential applications for the, to be created, Dutch governmental ontology in a central registry context. The main purposes of this registry are to make information more accessible and to allow for such information to be used to shape policy, to provide transparency and information to the data subject about a systems purpose and security measures, to identify relevant data sets, and allow for queries to be sent to the relevant location, and to provide data subjects with a more easily accessible way to both determine where their data is processed and to exercise their GDPR data ownership related rights. The registry, void of personal data, cannot result in a direct risk to the privacy and security of the data subject. However, exposing excess system and project data could invite targeted cyber-attacks, resulting in an indirect risk.

Section 8.3 describes how the, to be created, Dutch governmental ontology could be used to limit the amount of data being exposed, to what is required for an individual's informational need. This has been illustrated through three main use cases present in any system, which is the use case of identifying any specific entity through a search system, the use case of a more detailed overview of the identified entity in combination with any functions required, and the use case of aggregational statistics where information is related to a larger number of entities. In all of these use cases, the level of information returned to the user has been limited to ensure compliance with the principle of data minimalization and privacy and security by design and by default.

Section 8.4 presents a broad picture of the information flow in governmental healthcare systems realized through the proposed FAIR-based framework. This section outlines the system's federated data approach, ensuring that data remains under the control of the processing entity. This is accomplished through providing access to data, instead of providing the data set files themselves. This ensures that the organization responsible for data processing is always in complete control over their data, including having the sole responsibility of ensuring that security measures are implemented and that automated and manual review of access to data is properly executed. This section also elaborates on a backup solution compatible with maintaining full control, as having backups is still a critical requirement for any healthcare system.

Section 8.5 provides a detailed system design of a healthcare system, based on the requirements and specifications of such a system, as specified In Section 7. This design has been constructed in such a way as to adhere to the principles of data minimalization and privacy and security by design and by default.

Section 8.6 describes the governance process of the development and deployment of FAIR-based projects and systems, which are constructed based on the implementation described in Section 8.5. The proposed FAIR-based framework intends to make systems themselves interoperable and reusable, which is achieved through the development of a common base applicable to any system and to allow for the deployment and further development of any previously created system. This approach can be divided into three main situations, covering the development of any possible FAIR-based project, with these situations being based on the existence of any comparable project. In the first situation, no such comparable project exists and development would begin based on a previously created default project including all features described in Section 8.5. In the second situation, there exists a more comparable but not fully suitable project. As this project is itself based on the default project, it already includes all features described in Section 8.5, and would reduce development time based on the project-specific features and implementations that have already been implemented. In the third situation, there exists a fully suitable project, which can be deployed immediately as long as the required organizational measures are implemented. These situations would apply in a federated approach, where a different clinic, GGD, etc. wants to make use of a system that is already in use at another location. In all situations, project-specific features and implementations can be more easily implemented based on the fact that they all share a common base on which these features and implementations have been created, which ensures that even when no such comparable project exists, development would still be able to benefit from other projects.

Section 8.7 describes security vulnerabilities that are still present in the proposed FAIR-based framework, which has been included to prevent the notion that a system is completely secure just because it has made use of this framework. While these vulnerabilities can all be prevented, they are important aspects to consider. These vulnerabilities have been divided into two groups, with the first group describing the danger of network intrusions, which are reliant on the implementation of security measures at the facility instead of the security measures present in the FAIR-based system. The second group described vulnerabilities related to an incorrect implementation of the FAIR-based system, which is likely to appear given the need for such an approach to be highly customizable. The primary danger of this is that while the FAIR-based system can effectively limit data access, this is ultimately reliant on how the organization itself has limited it. If such access would be specified beyond the informational need of the user, it risks repeating the problems present in CoronIT and HPZone, with efficiency and speed being valued over security.

8.1 Dutch Governmental Ontology

The proposed FAIR-based framework should be supported by either one or multiple ontologies, operating on three different layers. The first layer is the organizational layer, which aims to create an effective and clear overview of the various organizations in the Dutch government. While such an overview is likely to exist, at least in certain parts, the main aim of this (part of the) ontology is to standardize this overview and show the relations between the various organizations. Additional information can then be added to further improve the value of this approach, with this information being able to be integrated into the rest of the framework, as well as to more easily compare different organizations with figures that are currently separated over a variety of different sources. The information in this layer has no risks to privacy, being information that is publicly accessible and in service of the Dutch Society. This will be further discussed in Section 8.1.1.

The second layer is the Project layer, which aims to create an effective and clear overview of projects used by the Dutch government, as well as showing which organizations are linked to each project, either through its use or the development of any given project. This layer can significantly aid the aim of transparency in the Dutch government, with individuals being able to clearly understand the purpose of projects, the storage duration, and the potential impact this could have on any given data project. By adding additional information and links to for example its open-source development location, the project tender, and the DPIA, it allows for individuals to more easily supervise the actions of the government as well as to participate in the design of systems, benefiting from a wisdom of the crowd approach. While the information in this layer also poses no direct risk to the privacy and security of the data subject, it may be warranted to limit the amount of information that is made available to the general public as such information may be used to invite targeted cyber-attacks on systems presenting a combination of large volume/high sensitivity data with no/weaker security measures. This will be further discussed in Section 8.1.2.

The third layer is the projects variable layer, which aims to create a clear and effective overview of the specific data usage in each project. In the context of the data subject, the most important information is the specific data variables that are being processed, the form in which this information is exposed to individuals, and the general sensitivity of the variable. In the context of system design, this ontology layer has a different purpose, being used to shape system design and ensure that concepts are consistent over all projects and organizations in the Dutch government. In the context of organizations and data scientists, this ontology layer creates a clear overview of which data is collected and processed in each organization, which allows for the identification of relevant projects, as well as risk assessment. This will be further discussed in Section 8.1.3

Creating the ontology itself, which this thesis has called the Dutch governmental ontology, was ultimately beyond the scope of this thesis and corresponding investigation. Instead, this section provides both the content and design with which such an ontology could be created by the Dutch government in the future. Previously created resources that may prove useful for the creation of this ontology have also been identified and included in this section, to further allow for the creation of this ontology. The RDF triple relations in this section, as divided into Predicate, object, and explanations, are not an exhaustive list of all possible relations but should instead serve as examples towards which information could be included in the ontology.

8.1.1 Organizational Layer

The organizational layer of the Dutch government already has well-established FAIR ontology which clearly describes all organizations falling under the Dutch government, as well as the relations between them and various additional relevant information. This is described through the "Thesauri en Ontologieën voor Overheidsinformatie", which was finalized on the 6th of December 2022 [205]. An overview of this implementation is depicted in Figure 11 [206], with the class diagram depicted in Figure 12 [207].

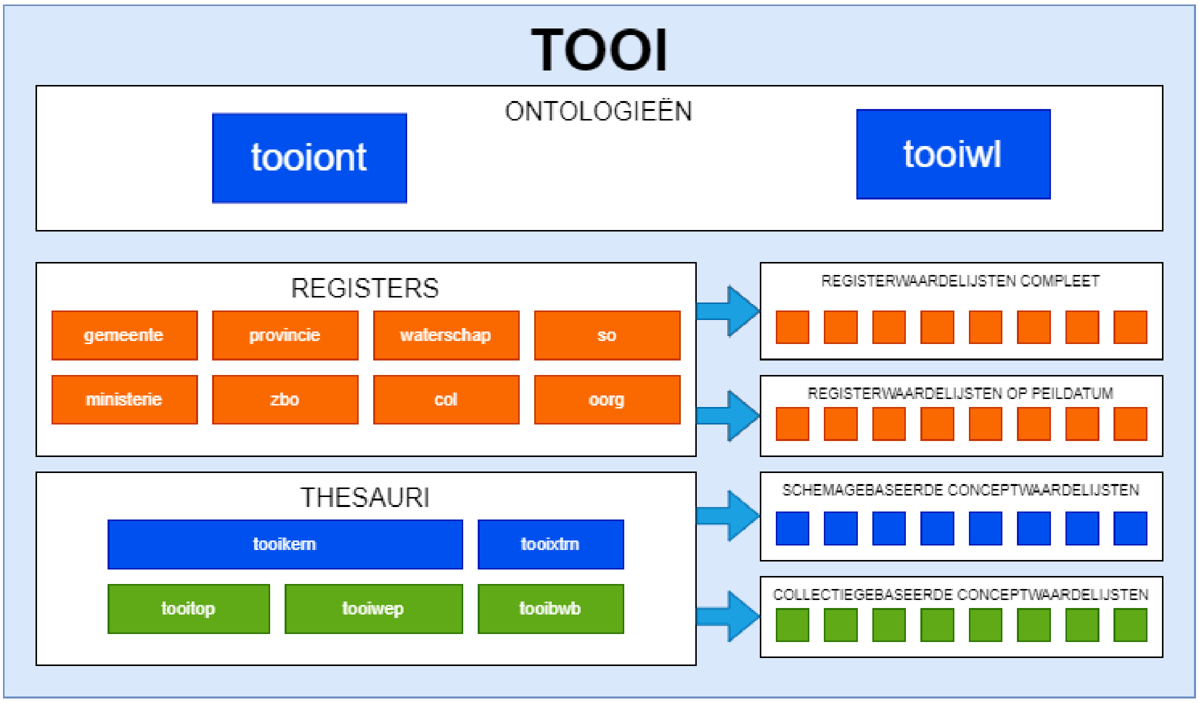

TOOI
ONTOLOGIEËN
tooiont
tooiwl
REGISTERS
gemeente
provincie
waterschap
so
ministerie
zbo
col
oorg
REGISTERWAARDELIJSTEN COMPLEET
REGISTERWAARDELIJSTEN OP PEILDATUM
THESAURI
tooikern
tooixtrn
tooitop
tooiwep
tooibwb
SCHEMAGEBASEERDE CONCEPTWAARDELIJSTEN
COLLECTIEGEBASEERDE CONCEPTWAARDELIJSTEN

Figure 11 - Thesauri en Ontologieën voor Overheidsinformatie

Given that this implementation already fulfills the requirements that this framework would require of an ontology, this thesis proposes to make use of this ontology, in the creation of the overall Dutch Governmental Ontology, to support the rest of the framework. In addition to this, the ontology itself is already publicly accessible, structured, and searchable [208]. Further improvements can be made to both expand the use case of this ontology, however, as it was designed to provide contact information for each governmental entity, instead of the broader use case presented in this thesis. These improvements are related to improving the search functionality, including additional relations, connecting and making use of the relations that currently exist, and offerings ways to make queries into this large repository of information.

The search functionality can be improved by expanding the search engine to be able to search both the content of each entity, as well as the inclusion of commonly associated names. For example, 'Rijnsburg' is a village that's part of the municipality of 'Katwijk', which is included in the information of the entity 'Katwijk' but is unable to be identified by this search engine. Similarly, any organization can only be identified directly by name, with the system being unable to present a list of any organization located in 'the Hague' for example, as the search engine does not take into account any included information for any entity.

Current entities also lack certain expected links between classes, while each individual class has been included. For example, all municipalities and all provinces have been included in this registry. As each municipality is part of a province, this relation should be included, but isn't at present [209]

[210]. Other information that would make use of this relations has also not been included. For example, while municipalities list the size, population, and population per area the provinces that these municipalities are a part of do not list this information. As the information of each province is a combination of the numbers of each of its corresponding municipality, this is simplistic to include.

The interaction with the registry could also be expanded to introduce queries and filters for additional information, based on data that is currently already present in the system. For example, each municipality has a list of political parties, with the number of seats and the total size of the municipality council, but the interface does not offer any way to compare this information to other municipalities. Neither is a link made that would show the presence of parties across all municipalities they have a presence in. Although this and the previous examples are ideas that could be included in the future but likely haven't due to the recent implementation of this system.

In addition to these suggested improvements, this thesis proposes the inclusion of additional variables which would offer significant value in transparency but have not been included. These variables are the number of employees at each organization and the budget of each organization, either historical, current, or proposed figures. While this information is generally public, this information is currently only located in either the yearly "Miljoenennota" or in the annual or quarterly reports of each organization, making it time-consuming to compare different organizations. The inclusion of such variables into (part of) the ontology would make this information easily accessible to the general public.

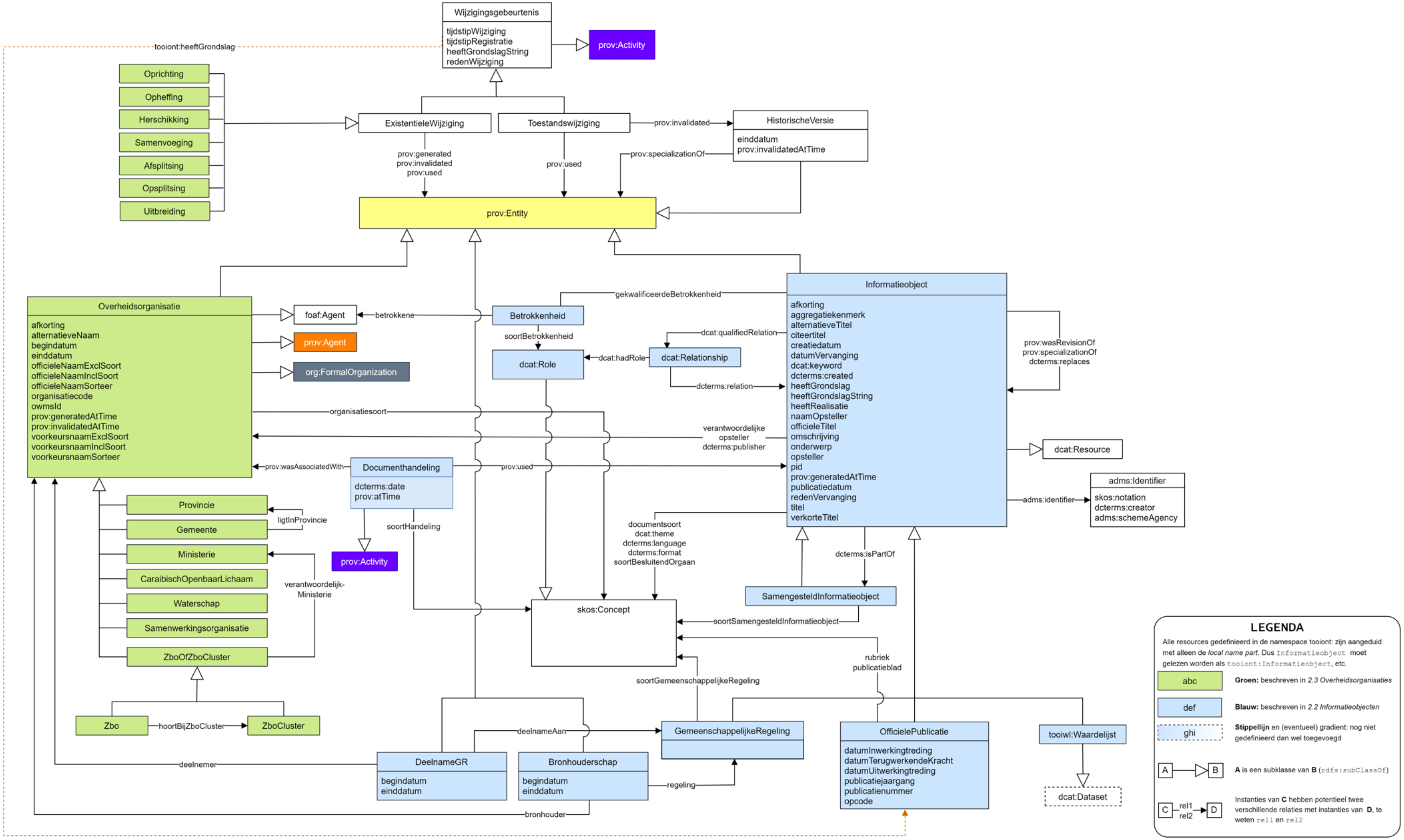
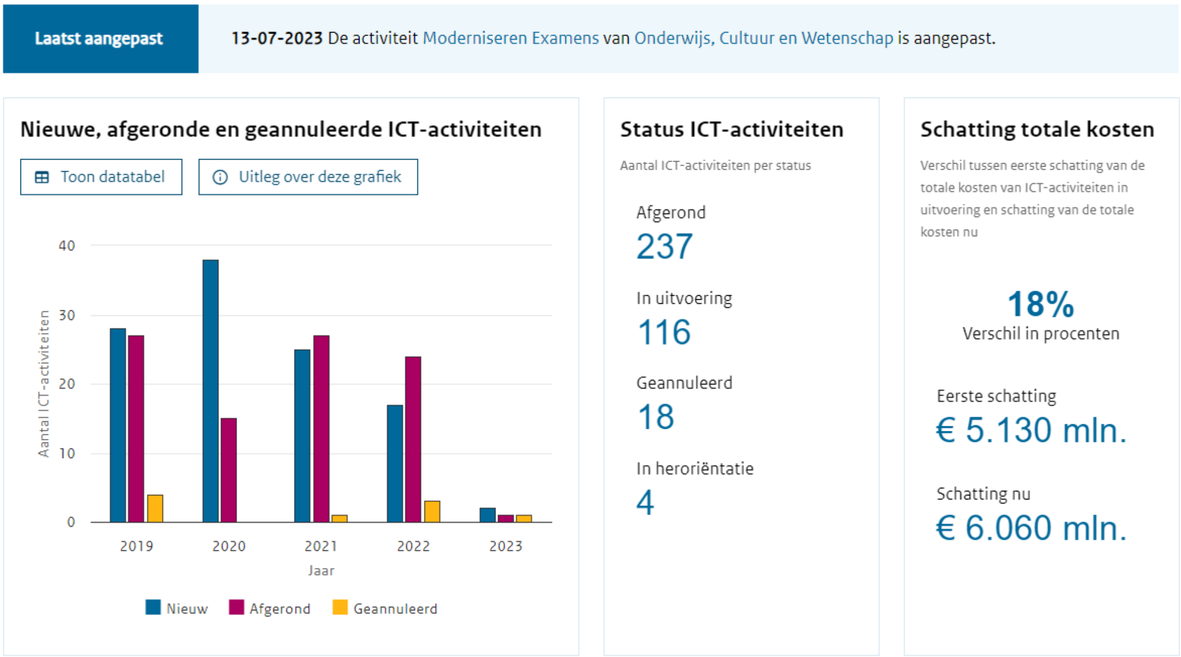


Figure 12 - Class Diagram TOOI

8.1.2 Project Layer

A project layer ontology currently does not exist, as far as was able to be found by this Thesis. The Dutch government did recently start providing some level of oversight for projects, in the form of a yearly report [211] [212] [213] [214] and the recently created “Rijks ICT-dashboard” depicted in Figure 13 [215], but these do not meet the requirements for the proposed FAIR-based Framework. Both of these forms of reporting only include large ICT projects, which are projects with an ICT component of at least 5 million euros and fall under the mandate of the AcICT, as previously discussed in Section 4.2. The information that is provided by both these sources is limited, with information being limited to the relevant ministry, the name of the project, its status, various financial figures, a description of the project, relevant dates, the type of development, and various information related to the executing party. This information is not built based on a FAIR ontology and is not connected to the previous layer in any way, which are important requirements for the proposed FAIR-based framework. Additionally, the dashboard is not complete, with information either not being present in specific projects or projects not being listed at all. For example, CoronIT is not included either in the Rijks ICT-dashboard or the yearly reports, while meeting the requirements based on its initial tender and the significant expenditure after the project was awarded to GGD GHOR.



Laatst aangepast **13-07-2023** De activiteit Moderniseren Examens van Onderwijs, Cultuur en Wetenschap is aangepast.

Nieuwe, afgeronde en geannuleerde ICT-activiteiten

Status ICT-activiteiten

Aantal ICT-activiteiten per status

Afgerond
237

In uitvoering
116

Geannuleerd
18

In heroriëntatie
4

Schatting totale kosten

Verschil tussen eerste schatting van de totale kosten van ICT-activiteiten in uitvoering en schatting van de totale kosten nu

18%
Verschil in procenten

Eerste schatting
€ 5.130 mln.

Schatting nu
€ 6.060 mln.

Figure 13 - Rijks ICT-dashboard

To meet the requirements of the proposed FAIR-based framework this approach will need to be expanded to cover all projects instead of being limited to recent projects that also qualify as large ICT projects. In the context of this thesis, this approach should be focused on projects involved with the processing of personal data related to data subjects. This would require all older projects or at least all projects that either still actively process personal data or projects that still contain personal data as specified by their storage policy duration, to have their relevant information recorded and be publicly accessible. In the future, the structured approach could also be applied to ICT projects that do not meet these conditions, but these are not relevant to this thesis.

Additionally, a distinction can be made between projects that are currently in development and are not actively processing personal data yet, projects that are currently in use, and projects that are currently in use but are also under current development. In all instances, the data related to these projects can be of value, but this value is not shared between all groups. For example, all projects would benefit from a reference to the location of the open-source development, either to address or improve certain aspects. But from the perspective of civilians without this technical background, the most important projects are the projects that are currently processing their personal information.

To accomplish this, the current data, and the additional variables that this thesis proposes, will need to be structured into an ontology. This ontology also needs to be connected to the organizational layer to be able to create an overview of projects belonging to a certain organization, either through its development or its use. Another condition is that systems can either use a federated approach where each organization has its own separated instance of the system and data within it, or that a centralized approach is used, with each organization having access to a central system containing all data from all organizations. An overview of these variables, structured in RDF triples is depicted in Table 11, which can be used to create the actual ontology in the future.

Predicate	Object	Explanation
containsSensitiveInfo	xsd:boolean	The project contains sensitive information, based on the data variables in the next section
containsMedicalInfo	xsd:boolean	The project contains medical information, based on the data variables in the next section
hasProjectName	xsd:annotation	The name of the project/system
hasCategory	Entity	The category or multiple categories to which the project belongs. E.g., Covid, Medical
hasGoalAreas	Entity	The areas of goals this project is an improvement on. E.g., efficiency, improved service, etc
hasMinistry	Entity	The ministry under which the project is organized
hasMinister	Entity	The minister that approved the project
hasDevelopmentStartDate	xsd:date	The start date of project development
hasDevelopmentProjectedEndDate	xsd:date	The projected end date of project development
hasDevelopmentRevisedEndDate	xsd:date	The revised end date of project development
hasDevelopmentDuration	xsd:integer	The project development has taken X number of years/days based on the current date if the project has not finished the initial development
hasDevelopmentDurationProjected	xsd:integer	The project development is projected to take X number of days based on the revised end date
hasDevelopmentType	Entity	The type of project development e.g., Waterfall, agile
hasProcessingStartDate	xsd:date	The start date of data processing
hasProcessingEndDate	xsd:date	The end date of data processing
hasDescription	xsd:date	The description of the project/system
hasPurpose	xsd:annotation	The description of the purpose of the project
hasDevelopmentCosts	xsd:integer	The development cost of the project
hasReferenceToOpenSource	xsd:annotation	A link to the open-source development of the project
hasReferenceToDPIA	xsd:annotation	A link to the Data Protection Impact Assessment(s) of the project
hasReferenceToAcICTAdvise	xsd:annotation	A link to an assessment of the AcICT
hasReferenceToDashboard	xsd:annotation	A link to the overview on the Rijks ICT-dashboard, as not all information needs to be presented
hasReferenceToPubicTender	xsd:annotation	A link to the project tender on TenderNed

HasPublicTender	Xsd:boolean	The project was executed through the release of a public tender on TenderNed
hasStorageDuration	xsd:integer	The number of days/years that data is stored by the organization
hasNumberOfUsers	xsd:integer	The number of individuals with access to the system
hasUsersType	xsd:annotation	The description of the individuals processing data in the system
hasProfessionalUsers	xsd:boolean	The system is only used by trained professionals e.g., medical staff
hasNumberOfDataSubjects	xsd:integer	The number of data subjects that have been stored in this system
hasRegionOfOperation	Entity	The region the system operates in, municipalities, provinces, etc
Has2FA	xsd:boolean	2-Factor authentication is used to access the system
hasSystemAdress	xsd:annotation	The link to which API requests can be made to interact with the system
isInActiveUse	xsd:boolean	The system is in active use
isDevelopedBy	Entity	The organization responsible for developing the project
isDevelopedInternally	xsd:boolean	The organization developing the project is the same organization processing the data
isDevelopedExternally	xsd:boolean	The organization developing the project is different from the organization processing the data
isDevelopedCommercially	xsd:boolean	The organization developing the project is a commercial party that bid on the project
IsDevelopedWithFAIRFramework	xsd:boolean	The project is developed based on the proposed FAIR-based framework
isDerivedFrom	Entity	The project is developed based on a different project as a starting point
isLargeICTProject	xsd:boolean	The project has a development cost of 5.000.000 euros or more, based on the development cost value
isDevelopedOpenSource	xsd:boolean	The project development is open source
isInfoUpdatedDate	xsd:date	The most recent date changes have been made to this overview of project data
isUpdatedDate	xsd:date	The most recent date changes have been made to the project itself
IsNationalSystem	xsd:boolean	The system operates on a national scale
hasVersionNumber	xsd:annotation	The current version number of the project
hasDPIA	xsd:boolean	The project has a DPIA
hasAcICTAssessment	xsd:boolean	The project has an assessment by the AcICT
isUsedBy	Entity	The project is used by X organization
isFederatedAccess	xsd:boolean	The project uses a federated approach with each organization running its own instance
isCentralAccess	xsd:boolean	The project uses a central approach with each organization connecting to a central database

Table 11 - Overview of RDF triples of the Project Layer

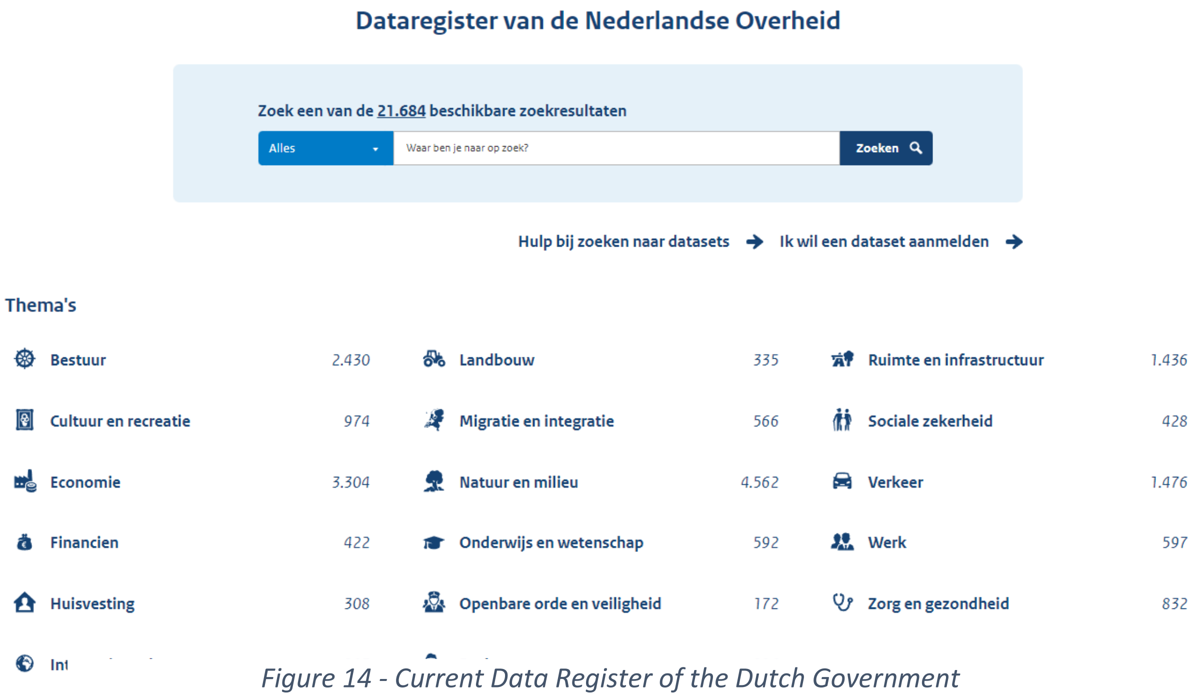
8.1.3 Project Variables Layer

The project variable layer is the lowest layer of the proposed Dutch ontology within the FAIR-based framework, being directly connected to the previous Project Layer, which is itself structured under the Organizational Layer. In the context of the overall ontology, the purpose of this layer is to regulate the usage of each specific variable used by any governmental healthcare organization in the Netherlands. This approach may also be of significant value to organizations beyond this specific sector, the government, or even the Netherlands itself, but this falls outside of the scope of this thesis.

In the context of a specific project, the purpose of this layer is to act as a repository for the variables used in this project. This layer can be divided into two main groups, each serving a specific purpose. The first group focuses on the current reasons why FAIR principles are employed in data processing. Its objective being to establish a common vocabulary and shared understanding of data elements, their relationships, and semantics. By standardizing these aspects across all organizations working with the data, better data interoperability and analysis can be achieved.

The second group aims to incorporate elements promoting GDPR compliance, as well as the methods used to achieve this, into an ontology. This involves extending the common vocabulary with information that aids in meeting the requirements of the General Data Protection Regulation. The emphasis here is on enhancing technical and organizational security, particularly in safeguarding against internal threats from employees. This includes increasing risk awareness and assisting in the design of systems to mitigate potential privacy breaches. This is also achieved through the connection to the previous Project Layer, which includes additional information relevant to GDPR compliance.

The Dutch government has previously created a data register for various datasets created and maintained by the Dutch government, as illustrated in Figure 14 [216]. Although these datasets are supported by a FAIR ontology, this support is limited to describing the data set itself instead of the variables that are contained within it. In his way, this ontology operates on a similar level as the project layer ontology, which describes the overall project, and all relevant information, but does



Dataregister van de Nederlandse Overheid

Zoek een van de 21.684 beschikbare zoekresultaten

Alles
Waar ben je naar op zoek?
Zoeken
Hulp bij zoeken naar datasets
Ik wil een dataset aanmelden

Thema's

Bestuur	2.430	Landbouw	335	Ruimte en infrastructuur	1.436
Cultuur en recreatie	974	Migratie en integratie	566	Sociale zekerheid	428
Economie	3.304	Natuur en milieu	4.562	Verkeer	1.476
Financien	422	Onderwijs en wetenschap	592	Werk	597
Huisvesting	308	Openbare orde en veiligheid	172	Zorg en gezondheid	832
Int					

Figure 14 - Current Data Register of the Dutch Government

not describe the data within it. As a result of this, there are no references to any ontologies used to describe data variables and relations, limiting its value in the context of standardizing data usage, enabling scientific research, or providing transparency. Due to this, this implementation is not useable for the proposed FAIR-based framework.

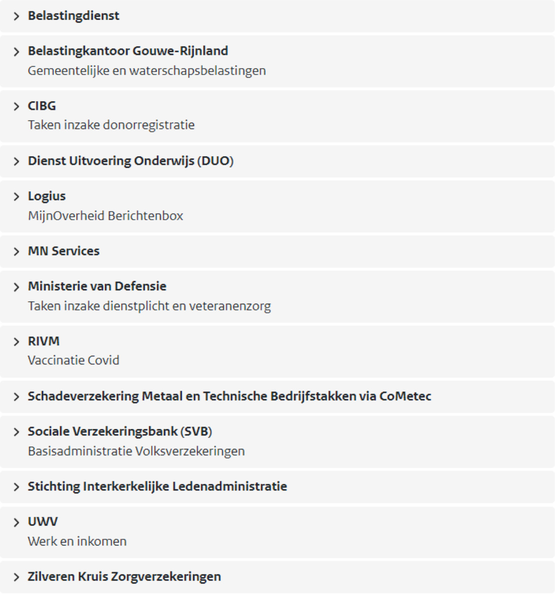
Instead, a new implementation is required that makes use of an ontology that describes data variables and their relations, to effectively standardize data usage. In addition to this, this thesis proposes to extend this with a proposed GDPR-related ontology, providing information and standardizing the usage of data within organizations, for example in the form of derivatives. The content and design of which have been included here to aid the creation of the Dutch Governmental Ontology in the future.

8.1.3.1 Default FAIR Ontology Usage

In the context of medical data, the Dutch government has already (partially) adopted an existing ontology, SNOMED CT, as part of the Dutch "Eenheid van taal" which was previously discussed in Section 6.4 [217]. The Ministry of public health is the Dutch license holder for this ontology, with Nictiz being responsible for developing and maintaining the Dutch version of SNOMED CT [218]. This project was started at the end of 2017 [219], by first translating (part of) the existing English-based ontology into Dutch. The translation of all Dutch relevant terms was completed in 2021 [220]. Given that this ontology is already well established, with a Dutch version, and at least partial adoption by the Dutch government, this thesis proposes that this ontology should be used as the medical terminology of the data processed in Dutch governmental healthcare systems.

The previously stated medical data is connected to a specific individual, based on the individual's BSN, as required by Dutch law for any system processing medical data. Using this BSN, a unique number given to any Dutch citizen, or an individual working or studying in the Netherlands for a period of time can be uniquely identified when interacting with the Dutch government [221]. This specific individual then has certain non-medical properties that should also be included in an ontology such as their name (first, initials, middle, last), and location of residence.

A source for this information is the Basisregistratie Personen [222], depicted in Figure 60, which has structured this information, but does not make use of any kind of ontology. However, due to the simplicity of this information, it could easily be converted into an ontology that meets the requirements of this approach, or the information could be stored following the FOAF ontology [223]. This source of information also sees limited use across Dutch organizations, as depicted in Figure 15 [224], with for example the GGD not being included in this list. Instead, organizations choose to store this information themselves, increasing the value of having an ontology standardizing this information even further.



› **Belastingdienst**

› **Belastingkantoor Gouwe-Rijnland**
Gemeentelijke en waterschapsbelastingen

› **CIBG**
Taken inzake donorregistratie

› **Dienst Uitvoering Onderwijs (DUO)**

› **Logius**
MijnOverheid Berichtenbox

› **MN Services**

› **Ministerie van Defensie**
Taken inzake dienstplicht en veteranenzorg

› **RIVM**
Vaccinatie Covid

› **Schadeverzekering Metaal en Technische Bedrijfstakken via CoMetec**

› **Sociale Verzekeringsbank (SVB)**
Basisadministratie Volksverzekeringen

› **Stichting Interkerkelijke Ledenadministratie**

› **UWV**
Werk en inkomen

› **Zilveren Kruis Zorgverzekeringen**

Figure 15 - Overview of the sharing of information from the BRP

Regarding the location of residence, this information could make use of an ontology focussed on locational data. As well as being interconnected with the organizational layer of the ontology, which would connect for example municipalities or provinces with their information and relations stored in that ontology. The information contained in this ontology should enable any kind of investigation concerning locational data. This, in addition to the location of residence, could also be related to locations that individuals visited in the context of contact tracing investigations. A comprehensive ontology for location data would include the following information:

- the latitude and longitude of the specific location, based on the other information
- The country, province, and municipality of the specific location
- The street and house number of the specific location

8.1.3.2 Ontology Extension to Incorporate GDPR Compliance

In the context of the GDPR, the focus of this (part of the) to be created ontology will be on aiding the technical and organizational measures implemented to ensure privacy and security by design and data minimalization. While FAIR and scientific data analysis assumes to be working with complete data, this extent of data exposure is often not required in the context of the primary healthcare process. In this process, information should be limited to what is required for fulfilling any specific task, based on the nature of the data and the expertise of the employee accessing the data. This ontology aims to address this by providing standard practices for each different variable, which can then be applied in specific situations, to limit the amount of data exposed while still being able to fulfill any specific task. The most likely usage for these will be as recommendations used during the construction of systems, although any deviation from a recommendation will likely need to be justified. Ensuring that even though this system isn’t a legal requirement, it has an increased amount of usage.

This, to be created, GDPR-focused ontology has been divided into three different Tables. The first table, Table 12, provides a general overview of the way that a specific data variable functions in a system. First of all, it contains basic GDPR-related information in the sensitivity of the variable, and its classification as common or special personal data, with another addition for criminal personal data, which would not be used frequently, but still has value in some healthcare systems. It finally contains a classifier for if the information is legally identifying, which is often the primary key used to identify any specific individual, such as our BSN. The rest of the information describes how the variable should be used in the system and the organization, which as previously mentioned are most likely to be recommendations. The actual usage of these is most logically structured depending on the specific use case and will be discussed in later sections of this thesis.

Predicate	Object	Explanation
hasSensitivity	Entity	The sensitivity of the variable e.g. None, Low, Medium, High
hasDefaultLevelOfUsers	Entity	The default level of expertise of users that would normally have access to the variable. The entity would refer to a type of employee, such as a medical professional, call center employee, etc. Each of those entities could add variables used to describe them
hasDefaultLevelOfInfo	Entity	The default level of information of the variable, as described in Table Table 13
hasDefaultLevelOfInteraction	Entity	The default level of interaction of the variable, as described in Table 13. For example, the default level of interaction could be to never expose the variable but instead use it as a function.
hasPublicDescription	xsd:annotation	A clear description of the variable, that can be understood by the general public. For example, in the context of a data subject request or included in a dashboard
hasProjectUsageDescription	xsd:annotation	A clear description of how the variable is used in the system, clearly separating different use cases and different levels of information in those use cases. For example, in the context of a data subject request or included in a dashboard
isCommonPersonalData	xsd:boolean	The data variable is not a special category of personal data
isCriminalPersonalData	xsd:boolean	The data variable is related to criminal convictions and offenses or related security
isSpecialPersonalData	xsd:boolean	The data variable is a special category of personal data e.g. racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data to uniquely identify a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation
isLegallyIdentifying	xsd:boolean	The data variable is data used to legally identify someone e.g., the BSN, which is a unique number attributed to every citizen of the Netherlands.

Table 12 - GDPR Related Data Variable Ontology

Table 13 provides an overview of ways in which data can be returned to the user, based on transformations of the original data value. This overview is certainly not comprehensive but serves only to demonstrate the general concept. When this concept is expanded, it could offer significant value in ensuring that organizations adhere to the principle of data minimalization. As the original data format is also predefined, based on the other aspect of the ontology that regulates the way that data is stored and the relations that this data has, it would offer significant value if code is created beforehand that could be used to apply each relevant transformation to the original data values.

Object	Description
CompleteValue	The data variable is returned to the user as the original value
ConvertToAgeYears	The data variable is converted to an integer age before being returned to the user. For example, a person's date of birth is converted to their age in years, which prevents the birth data from being exposed in the system
CalculateToGroup	The data variable is converted to another value before being returned to the user. For example, instead of a date of birth or age, the system would now return what it is used for. For example, age => 18 would indicate adult, and Blood pressure > 130/80 would indicate high blood pressure. This return condition is best used as an annotation instead of an actual function
NameInitials	The data variable is converted to another value before being returned to the user. For example, instead of an individual's name being Leendert van der Plas, it will be returned as L. van der Plas which would expose less information
LastCustomDigits	The data variable is limited to the final X digits before being returned to the user. The custom part refers to the ability to specify the number of digits instead of this being fixed. For example, instead of the full BSN being returned, it would now return the last X digits.
LastThreeDigits	The data variable is limited to the final 3 digits before being returned to the user. For example, instead of the full BSN being returned, it would now return the last 3 digits.

Table 13 - Level of Information Exposed of Data Variable

Table 14 provides an overview of ways in which the user would be able to interact with the data variable. It is important to note that this interaction does not have to involve the complete data variable, but instead could be limited based on the level of information discussed in the previous paragraph. Additionally, even after user interaction, the information that is returned could be limited and still not be the original and complete data variable. The interaction itself can be divided into four main groups, with the first group not using any interaction, the second group requiring some sort of interaction to expose information, the third group using a confirmation function to verify information instead of revealing it and the fourth group using the data for the function which it is to be used for, instead of exposing any information. Together with the previous table, this concept will be explained in more detail in Section 8.5.4.

Object	Description
Value	The data variable is returned to the user, based on the extent of information described in the previous table.
InteractionExpose	The data variable is returned to the user, after user interaction. This variable could be combined with other classes that reduce the amount of information that is returned
Confirmation	The data variable isn't returned at all but is instead used in the form of a confirmation function. The only value that is returned is a Boolean if the entered information matches the record information. This can be applied either to the full value, or any limitation of this value.
FunctionalValue	The data variable isn't returned at all but is instead used for the function that it is used for. For example, instead of displaying an e-mail address, the system would simply mail the individual without exposing the number.

Table 14 - Level of Interaction with Data Variable

8.2 Central Registry Implementation

The content and design of the to be created ontology, as discussed in the previous section, can be used for a variety of purposes, depending on the information need of the individual. This section will discuss four of these uses, with information being accessible through the creation of a central registry containing this information, although downloadable datasets can also be offered for investigations exceeding the capabilities of the current implementation of the registry. It should be noted that these are examples of the use of the described ontology data, and not an exhaustive list of all possible uses. As the registry does not need to contain any personal data, only the described ontology data itself, the creation or use of this registry is unable to result in any potential risk to GDPR compliance. Instead, it can improve GDPR compliance across all governmental projects due to an increased amount of transparency and supervision and gaining better control over data usage across organizations.

One potential obstacle to this approach is the fact that projects not based on the proposed FAIR-based framework would have to be added as well to create a truly comprehensive record of projects and systems used by organizations. While it would be relatively straightforward to add these projects, adding the information specified in the Dutch Governmental Ontology, would most likely require some form of governmental intervention. The required information is only accessible in full at each organization, which will require some number of manhours to collect, verify and fill in the required information. This problem is increased with legacy systems, for which these records may not be easily accessible or have been stored at all. Due to this, in combination with a potential disinterest in additional oversight, at least some organizations would likely oppose this approach and either delay or neglect following this process unless it is mandated by the Dutch government. Especially because such comprehensive registers do not appear to currently exist, even though control over data is an important focus area.

The first purpose of the central registry is to create an overview of projects and data usage across the Dutch government, which can be used to both identify similar projects as required in Section 8.6 and to make recommendations and adjustments based on identified patterns. For example, based on the purpose of the project and the connected data variables, similar projects can be identified at different organizations to consider if these projects could offer value in the current situation. Additionally, by using included information such as the use of 2FA, the creation of DPIAs, or any other relevant information and combining that with the data variables that are processed, significant risks can be identified. This can then be used to address these issues that would previously not be transparent to any organization but the one managing the system. This purpose is most valuable for supervisory organizations, larger organizations responsible for many different projects, and the larger government as a whole. Although there could also be significant value for technical individuals or news organizations, with reports being able to lead to adjustments being made to the privacy and security of records in those systems. This is discussed in more detail in Section 8.2.1

The second purpose of the central registry is to allow civilians to understand how specific systems process their data. This is a less complex version of the first purpose of the central registry, removing the need to run complicated queries and instead being focused on accessing singular projects. This overview offers project-specific details related to the overall project, as well as each data variable in detail, in addition to the purpose of each specific variable and the conditions under which that variable is made accessible to employees. This overview also contains some information that can be used to improve oversight, such as the inclusion of links to the created DPIA and a project's open-source development location. If this information has not been included, it could imply a lack of security in the system and therefore also offer some value. This is discussed in more detail in Section 8.2.2.

The third purpose of the central registry is to provide data scientists and other employees with a way to both identify relevant data sources and act as an interface in which queries can be sent to specific systems and results be returned. It is important to note that the registry functions solely as an interface and does not have access to any personal data. Instead, the registry stores the address of each system, allowing queries to be sent to any, or a combination of, specific systems. This is discussed in more detail in Section 8.2.3.

This aspect of the architecture is the most complex from both technical and organizational perspectives, as it requires both a common framework able to process these queries as well as shared variables following a shared ontology allowing variables to be processed the same in each system. Additionally, permission must be granted from the owner of the data source, with each data source potentially having different access conditions. The optimal scenario for this is a project within an organization like the GGDs, where the included variables and access conditions are similar, with the only variation being the region of data collection and the GGD organization responsible for processing it. Although this assumes that there are access conditions, which may not also be restrictive in these systems given that aggregational statistics would already prevent personally identifiable information from being exposed. Additionally, anonymized data sets could still be provided as a fallback method, to match the functionality of the previously mentioned Data register of the Dutch Government.

The fourth and final purpose of the central registry is to enable data subjects with a simple and automated way to exercise their GDPR data-ownership-related rights. This can be achieved using two different approaches, with the first approach focusing on aiding the exercising of data-ownership-related rights on singular systems and the second approach combining the information from multiple/all systems into one comprehensive overview to aid the exercising of data-ownership-related rights on a far larger scale. The first approach is discussed in more detail in Section 8.5.6, with the second approach being discussed in more detail in Section 8.2.4.

In the context of healthcare data, both approaches can be achieved on a technical level by using a person’s BSN as the identifier. The BSN is a unique identifier for all citizens in the Netherlands and is legally required when processing medical data. Additionally, the BSN already allows citizens to access their data by utilizing DigID, using it as a secure authorization system to verify user identity and request individual records.

Implementing this in other sectors might be more challenging, as these systems might not make use of the BSN and therefore prevent both the use of DigID as the secure authorization method and preventing data from being linked using a shared unique identifier. To address this, it would require either incorporating the use of BSN into systems that previously did not require it and jeopardizing the security and privacy of these systems, or implementing an alternative identifier that is itself linked to the BSN in some way. As the scope of this thesis is limited to the governmental healthcare sector, which does use the BSN in all instances, this will not be explored further but can be addressed through potential Future Work.

8.2.1 Regulatory Overview

This section presents various uses of the regulatory overview, with the primary objective being to ensure control over data usage across organizations. This control extends not only to the data that is being processed in each system, but also to how this data is processed, and what organizational measures and procedures have been implemented. The security and privacy of personal data may also depend on a variety of other factors, which can only be identified if this data is both recorded and able to be analyzed. This section will discuss the content and design of the to be created governmental ontology in the context of these objectives, providing various examples of how this data can be used to promote the success of governmental projects in both their results and level of regulatory compliance.

8.2.1.1 Overall ICT Implementation

Given that this approach would result in a comprehensive overview of the usage of data in systems across a large number of organizations, this data can be used to monitor the overall implementation of ICT in the Netherlands. This can be done across a variety of different variables, not limited to what has previously been discussed in this thesis, but for the purposes of this thesis, only a few examples will be provided. These are the data processing in organizations, the extent of open-source development, the extent of 2FA usage, the usage of DPIAs, and the extent of legacy vs. non-legacy systems.

The first example is the extent of data processing in organizations. Currently, organizations may often not have a comprehensive overview of the systems they use, how they interact, or what data they even process. This approach would be able to register these systems and define all of these relations, which can then be used in the future when any of these systems needs to be changed/updated. With a comprehensive overview of the data contained in systems, similar systems could be identified and any potential redundancy could be eliminated. It also becomes possible to do a risk analysis of various systems that would previously have been impossible due to a lack of information.

The second example is the usage of 2FA in organizations. 2-Factor-Authentication is a simple technical measure that can be taken to prevent most unauthorized access in systems. While the measure itself is simplistic and does not impact the overall usage of any system, most older systems would lack such a measure as they would have been designed in a time before such implementations became standard. Currently, it would be difficult to determine how many, and which of these, systems, either have or have not implemented such a measure. With this ontology and organizations registering this information, this would become a simple query with the ability to directly measure progress in this area. While this specific example is about 2FA, this variable can realistically be applied to any technical or organizational measure.

The third example is the usage of open-source development in projects, which is one of the aims of the Dutch government. Currently, the usage of open-source development is limited in scope, with each project that uses it being reported upon, but this should not be an assumption for the future. To be able to determine this in the future, it needs to be recorded and easily viewable, and comparable wherever projects are developed through open-source development.

The fourth example is the usage of DPIAs in organizations. While DPIAs are created by default in many instances, not all systems have a DPIA. Limited registers for the creation of DPIAs exist, but these are not comprehensive and would only contain information about created DPIAs in any specific organization or overreaching other organizations. DPIAs also do not have to be made public, requiring either a WOO request in the context of civilians or access to such a register in the context of the DPO of any such organization. As a result, there is no overview of the created DPIAs nor is there an overview of which organizations have a DPIA. This ontology would address this by registering DPIAs to the relevant project, providing both access to these DPIAs as well as registering which projects have and have not created one before processing personal data. As DPIAs are important documents, as described in Section 4.3.1, the identification of these projects as well as monitoring the overall usage would be of significant value to the overall state of ICT in the Netherlands.

The final example is the extent of legacy versus not legacy systems in use in each organization. Legacy systems, being designed before the GDPR with no requirements to adhere to modern standards, can pose a significant risk to the privacy and security of the data within them. Creating an overview of these would offer substantial value. The information can be used, in combination with other information, to prioritize projects that could be updated to meet these modern requirements. If at any point, the system is updated to exceed the conditions of a legacy system, queries could identify such projects and either prevent the usage of the system under these conditions or prioritize the modernization of this system even further.

8.2.1.2 Pattern Detection

With a large amount of information from a variety of different projects, it also becomes possible to identify patterns in the development and success of projects. This approach shares some similarities with the Rijks ICT-dashboard but would be broader in scope.

This information can be used to either address various issues with specific aspects that prevent the success of any given project, or it could be used to favor other combinations of aspects to ensure that any given project has the highest probability of success. It can also be used to compare different organizations and determine why some of these organizations have success while other organizations don’t, with the ability to implement changes on an organizational level.

Examples of these comparisons are to for example compare the success of different ministries, which is a comparison that is often used in other types of reports. This distinction is made in the yearly review of the AcICT, the Rijks ICT-Dashboard, and data breaches as reported by the AP. However, the troubling aspect with each of these data sources is that this comparison is not done on any level lower than the Ministries themselves, for example, each of the relevant organizations that fall under the responsibility of the ministries. There is also no combined analysis that takes into account multiple factors.

Other comparisons that could be made are to evaluate the success of projects alongside different categories, for example, projects that aim to increase “efficiency” or an “improved level of service”. Another comparison is the success of projects that are either internally or externally developed, to shape the development of governmental ICT in the future. Another comparison is the success rate of projects of various levels of budget, to determine what is the optimal size for a project and which types of projects would require additional measures to exceed. Another comparison is the success rate of the development of projects by any specific organization, which would indicate which organizations would be assigned projects in the future or where intervention would be required. All of these comparisons can also be combined, to evaluate this with additional conditions, for example, the efficiency of the external development of large-scale systems in the Ministry of VWS.

8.2.1.3 Compliance in specific projects

With a large amount of information from a variety of different projects, it also becomes possible to follow up on certain events, based on meeting or not meeting certain conditions. While there are many possible conditions for this, the exact conditions would best be determined by the Dutch government and the supervisory organizations. This thesis will provide some examples, however, to demonstrate the value of this approach These examples are the usage of system version information, the creation of DPIAs, and evaluations by the AcICT.

The first example of a condition leading to further action could be to signal when the system version of a system changes or data variables are added, as it is updated, without there being a corresponding update of the DPIA. There are various types of updates, not all of which should trigger such a signal, but this example will assume there to be a substantial change instead of a simplistic security update. When an update of this magnitude occurs, the DPIA should be updated to reflect the change in the processing of personal data. When no such update to the DPIA occurs, organizations should be ordered to either update the DPIA or provide the reasoning why they aren't. In the case study into the GGD, this exact scenario happened, with CoronIT being used for a completely different purpose, without resulting in any update to the DPIA.

The second example of a condition leading to further action could be to signal when projects that are not legacy projects proceed with development, or even being used to process personal data, without the creation of a DPIA. As the data variables themselves are included in the project, this information can be used to further evaluate if the creation of a DPIA would be warranted for the system. This signal would prevent systems processing sensitive personal data from operating without even an evaluation of the system's impact and security standards.

The third example of a condition leading to further action would be an evaluation by the AcICT. Depending on the nature of this advice, either being positive or suggesting various changes, the project could be examined at a later date to evaluate if these changes have been implemented to improve the system. When the status of a system changes, this could trigger an immediate investigation to determine if the required changes have been implemented.

8.2.2 Informational Overview

This section presents various properties of the registry in the context of its informational purpose, alongside a possible implementation depicted in Figure 16. This example has used the information described in the to be created governmental ontology, but given the extent of the information, and the interest of citizens, different overviews could be created. The exact implementation of the features and UI as presented here may therefore differ from an actual implementation. The purpose of this example is solely to illustrate the value of such a registry, with an actual implementation being able to improve upon this initial design in the future.

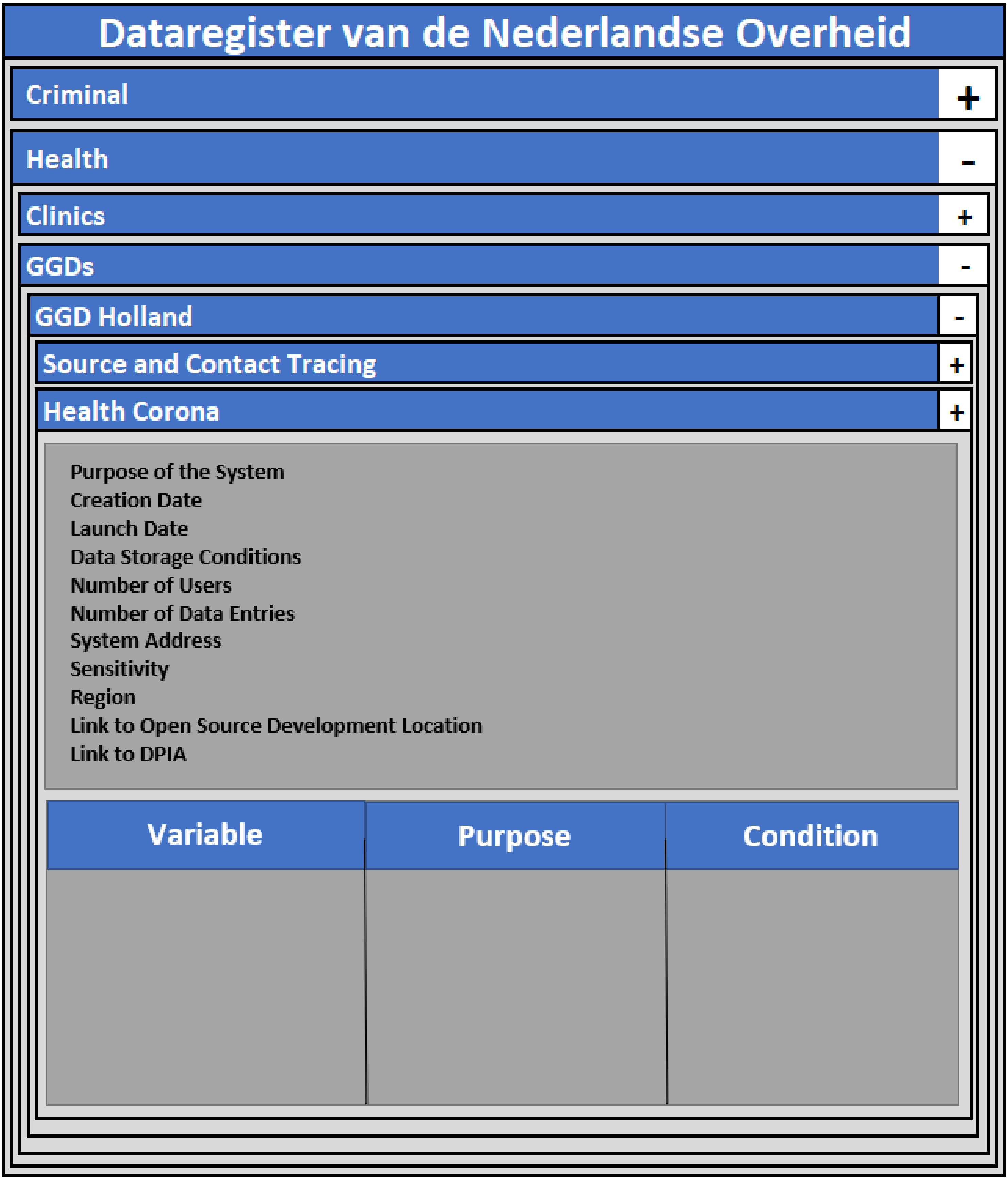
Instead of organizing this based on organizations, it may be more accessible for citizens if organizations and projects have been organized under the category of data they belong to. In this example, the categories are 'criminal' for criminal data and 'health' for health-related data. The 'criminal' category can be expanded using the plus sign in the top-right corner, while the 'health' category is shown expanded with various sub-categories that may be useful for certain types of data. In the context of this example, the sub-categories correspond to the type of facility providing healthcare, with the sub-category ‘clinics’ being expandable and the sub-category ‘GGDs’ being expanded and depicting information.

At a lower level, beneath categories and sub-categories, facilities corresponding to these values are displayed. In this example, these facilities are the GGD organizations in the Netherlands, with GGD Holland being the specific expanded facility. This section usually corresponds to a certain geographical location, with hospitals, clinics, and GGDs typically serving specific areas. However, this geographical information is also a part of the projects themselves.

At an even lower level, beneath facilities, various projects corresponding to these values are displayed. In this example, these are the project for ‘source and contact tracing’, which could be either HPZone, HPZone Lite, or the new GGD Contact application, and the project for ‘health corona’, which could be CoronIT or its replacement. Only the project for ‘health corona’ has been expanded. Upon expanding the project, detailed information related to the project is displayed. It should be noted however that some organizations may have hundreds of different projects and systems, such as the Dutch tax office, which would almost certainly require a filter functionality to identify the correct project(s).

At the lowest level, under the various projects, detailed information about specific projects is displayed. This displays aspects of both the Project layer and the Project variables layer. From the Project layer, this example includes the system's purpose, access reference, creation and discontinuation dates, storage duration and conditions, the presence of any sensitive information, region coverage, and whether the system uses the proposed FAIR-based framework. Documents like DPIAs can also be referenced here, which would remove the need to either contact an organization or submit a WOO-request to receive this information. From the Project variables layer, this example includes the specific data variables, their purpose in this specific system, and the conditions under which these can be accessed, which is discussed in more detail in Sections 8.3.1, 8.3.2, and 8.3.3.

Considering the potential size and usefulness of the central repository, a search system should be included to filter results based on various conditions. It is important that, unlike the current government register, this search system would be able to search the internal values of these projects, as well as related terms. For example, if an individual were to search for Covid systems, this search query should return all systems related to Covid, no matter if this information is contained in the purpose of the system, the name, the category, etc. Related information should also be displayed in search results, accounting for misspellings or different definitions.



Dataregister van de Nederlandse Overheid

Criminal +

Health -

Clinics +

GGDs -

GGD Holland -

Source and Contact Tracing +

Health Corona +

Purpose of the System
Creation Date
Launch Date
Data Storage Conditions
Number of Users
Number of Data Entries
System Address
Sensitivity
Region
Link to Open Source Development Location
Link to DPIA

Variable	Purpose	Condition

Figure 16 - Example implementation of the informational central registry

8.2.3 Data Science Overview

This section outlines various properties of the central registry in the context of its data science purpose, which involves accessing one or more datasets for aggregational statistical analysis. An example implementation of the registry, depicted in Figure 18, includes features aimed at enhancing its usefulness and increasing the likelihood of adoption. It is important to note that the features and UI presented here may differ in an actual implementation, but provide a solid basis for consideration. Any actual implementation should improve upon this proposed version.

A limitation of this approach is that it can guarantee security, as explained further in Section 8.3.3, but only works when datasets do not need to be combined. This approach can combine the results of two separate queries run on two separate datasets with the same information, but not the datasets themselves. Combining datasets would necessitate transferring data from one or both facilities to another location, which is beyond the scope of this thesis but explored in Section 11.3.1.3 of the future works section.

The first issue to address is granting each user access to specific databases. While users do not need the same credentials at every facility, they must have valid credentials for each facility, necessitating a system for managing this. Facilities must either process access requests from users or utilize groups of users to grant access.

This thesis proposes a system that combines these approaches, focusing on using user groups corresponding to each facility. User groups simplify the access provision process, as adding or removing users is no longer the responsibility of the organization managing the dataset. It is also unlikely that any external user would need different access conditions from a user group. If individual users require different access conditions, the same system can be employed, but it is not the intended protocol for this architecture. Relying on this method may indicate a problem at the facility requesting access due to misconfigured user groups.

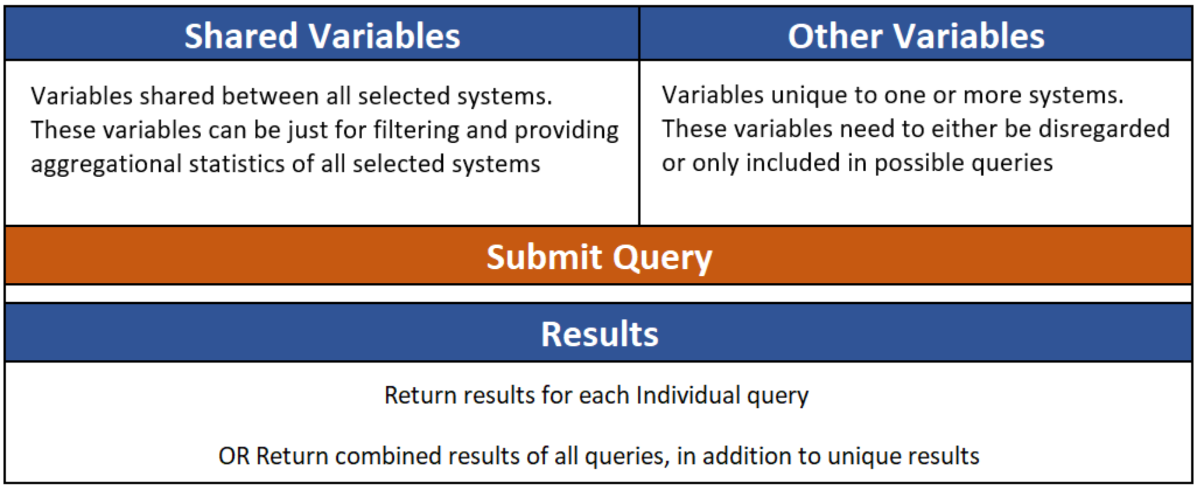
To illustrate how this system works, we first discuss user groups in the context of trusted third-party organizations, such as the CBS and RIVM, and then in the context of other organizations, such as research facilities, universities, and other organizations managing datasets referenced in the central registry. Trusted third-party organizations are organizations currently provided with access to data sets at various organizations in many instances, including in the case study discussed in the first part of this thesis.

Organizations managing datasets begin by sending a reference to a file defining access for user groups and specific individuals to the central registry, which is then accessed by the central registry on a schedule. This file contains information about user groups and specific individuals granted access, including the level of access, based on the dataset's nature and the organization type. The user group is managed by the organization provided with access, allowing them to add or remove users without individually sending changes to the organization granting access. Trusted third-party organizations, like the CBS and RIVM, are included in this file by default, while other organizations can be added based on the project or upon request. The level of access can be adjusted depending on the research type and the risk posed to the data subject, with the default access level balanced concerning this fact.

User credential validation is separate from the central registry and is managed by the user's organization, which would process the login request. Ideally, this process would use two-factor authentication (2FA) for validation. Users could then log in to the central registry using credentials from organizations such as CBS or RIVM, which are also validated with 2FA. After successful validation, users gain access to each dataset associated with their user group, subject to the specified access conditions. However, the data file only indicates a user's access level, as the data-providing organization is responsible for executing the actual query and denying it if the request is incorrect. The registry can only depict to the user, based on the information they have received, which data sets are accessible to the user or request access on behalf of the user.

Upon successful validation, users are presented with a screen similar to the example in Figure 18, constructed like the screen in Figure 16. The key difference is that users can either select datasets they have access to or request access to a specific dataset. In the context of requesting access, this would show all projects and data sets that meet a filter requirement while in the context of normal operations, this would most likely use a filter depicting only projects and data sets to which access has already been provided.

When only one dataset is chosen, all variables are shared by default, and the system performs similarly to the aggregation statistics system discussed in Section 8.3.3, although access conditions for external users may differ from those for internal users. When multiple datasets are selected, there may be differences in variables or user access to specific variables. This is visualized by dividing variables into two groups, 'shared' and 'other,' as shown in Figure 17. Since all queries are sent separately, 'other' variables can still be included and returned for the specific facility, but there may be issues with combining the data. For instance, if filtering is applied to a variable that does not exist or is inaccessible to the user, the query cannot be executed. This problem is exacerbated when datasets need to be interconnected, which is beyond the scope of this thesis and would be suitable for further study in a more comprehensive version of the proposed FAIR-based framework.



Shared Variables

Variables shared between all selected systems. These variables can be just for filtering and providing aggregational statistics of all selected systems

Other Variables

Variables unique to one or more systems. These variables need to either be disregarded or only included in possible queries

Submit Query

Results

Return results for each Individual query

OR Return combined results of all queries, in addition to unique results

Figure 17 - Example implementation of the data science central registry Query Implementation

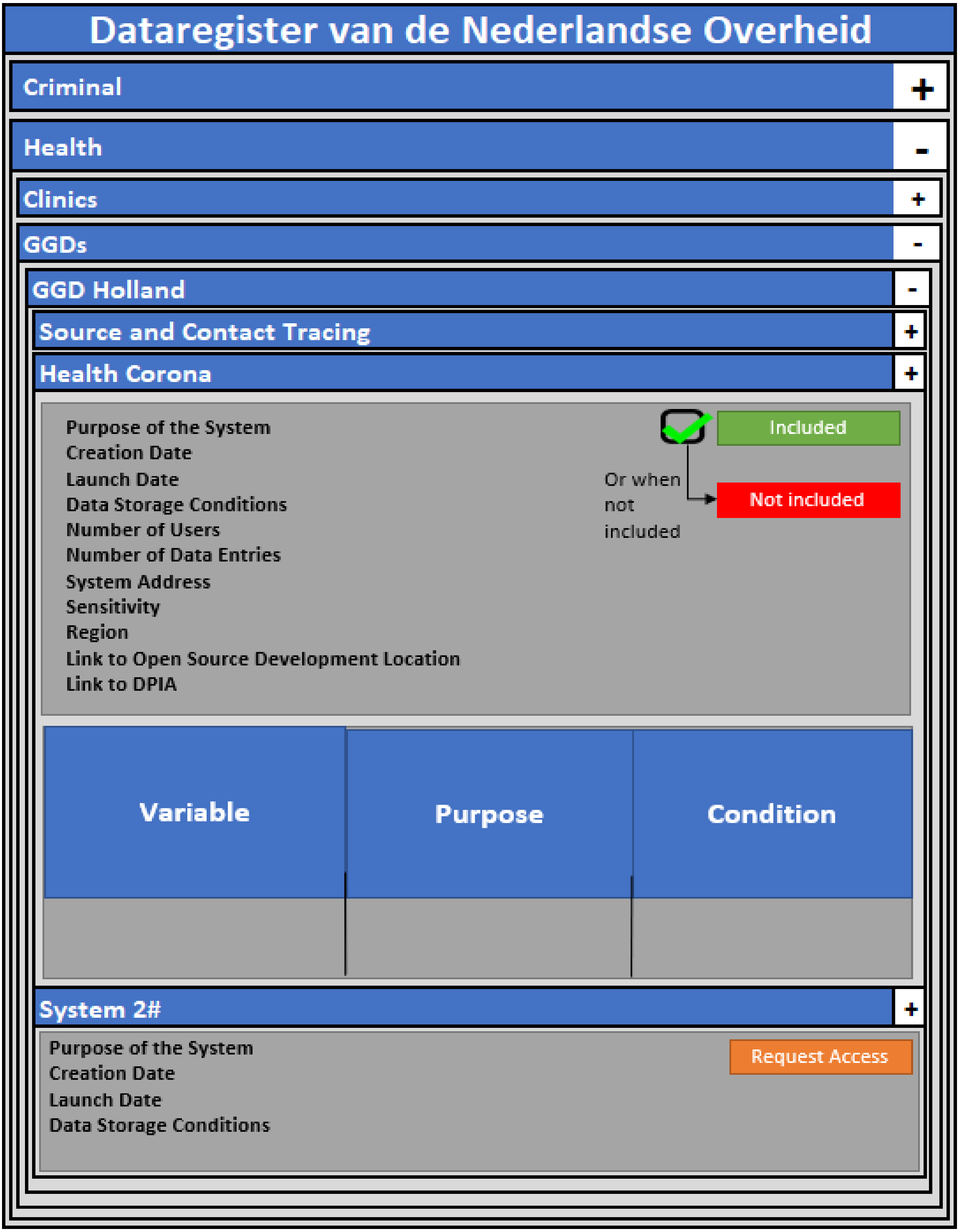


Figure 18 - Example implementation of the data science central registry

8.2.4 Exercising Data Ownership-Related Rights

This section outlines various properties of the central registry in the context of enabling data subjects to exercise their data ownership-related rights, as specified by the GDPR. While data subjects are always able to exercise their rights, this is often not an automated and easy process, requiring the data subject to contact an organization by e-mail, phone, letter, etc., and requesting one of these rights to be exercised. An example of a process that does meet these requirements is discussed in Section 8.5.6, which operates on the level of singular systems

However, initiating this process requires the data subject to know which organizations process and store data related to the data subject. Given the number of organizations that are part of the Dutch government, although limited by the geographical region the data subject lives in, determining this would currently take a significant amount of time and effort.

To address this, the central registry could be used to identify the processing of data related to the data subject across all projects and organizations that have been included at a single glance. To do this in a manner that respects the privacy and security of this information, as even knowing an individual has been included in certain systems could have significant implications, giving out this information should only be done to the data subject in question. The current method of verifying any person's identity online when interacting with governmental systems is DigID.

For its implementation in the central registry, the primary challenges include DigID's reliance on the inclusion of a BSN in the data set and the uncertainty surrounding DigID's pricing structure. These will be further explained alongside an example implementation of the registry

8.2.4.1 Potential obstacles

Considering the system's goal to function across all projects and organizations processing data related to data subjects, the pricing structure for DigID could be particularly problematic due to the high number of requests required to access information. If a single identity validation can be used to prove a person's identity for multiple systems simultaneously, DigID integration would be relatively simple and affordable, with a current cost of 13 cents being charged per use [225]. However, if each system requires separate identity validation through distinct DigID instances or if multiple systems necessitate separate identity validation through multiple DigID instances, the system's costs would likely hinder implementation. In the worst-case scenario, assuming hundreds of systems are added to the registry, costs could escalate to dozens of euros per use. If every individual in the Netherlands were to use this registry once, even with a relatively low number of systems (100), the total cost would exceed Logius' entire 2020 budget of 223 million euros [226]. While the actual cost to Logius of an additional request is likely to be substantially lower, with computational costs being only a part of their total budget, this example aims to show the sheer number of requests that would be made. Even if the cost per use can be significantly decreased, this would still result in a significant cost.

If the cost problem related to DigID occurs and cannot be addressed through changes at Logius, it may be addressed through the development of an authenticator similar to DigID. This could be done either by completely replacing the role of DigID or to use DigID to verify a person's identity to this alternative authenticator, which is then able to connect to all connected systems. However, the development of such a system might incur costs similar to those of creating DigID, given the high number of requests and stringent security requirements. As this system would be used to verify the identity of every citizen in the Netherlands and provide access to the most sensitive information about data subjects, security must be guaranteed.

An alternative implementation of this register in the context of data subject access could involve allowing users to access individual systems one at a time, with the central registry serving as a convenient access point. This approach might address the issue of numerous requests and the corresponding high costs associated with such a system, functioning similarly to the implementation described in Section 8.5.6. However, to be able to improve upon the current situations where it may not be clear which organizations process data of the data subject, this would require the central registry to know which systems contain information related to the data subject.

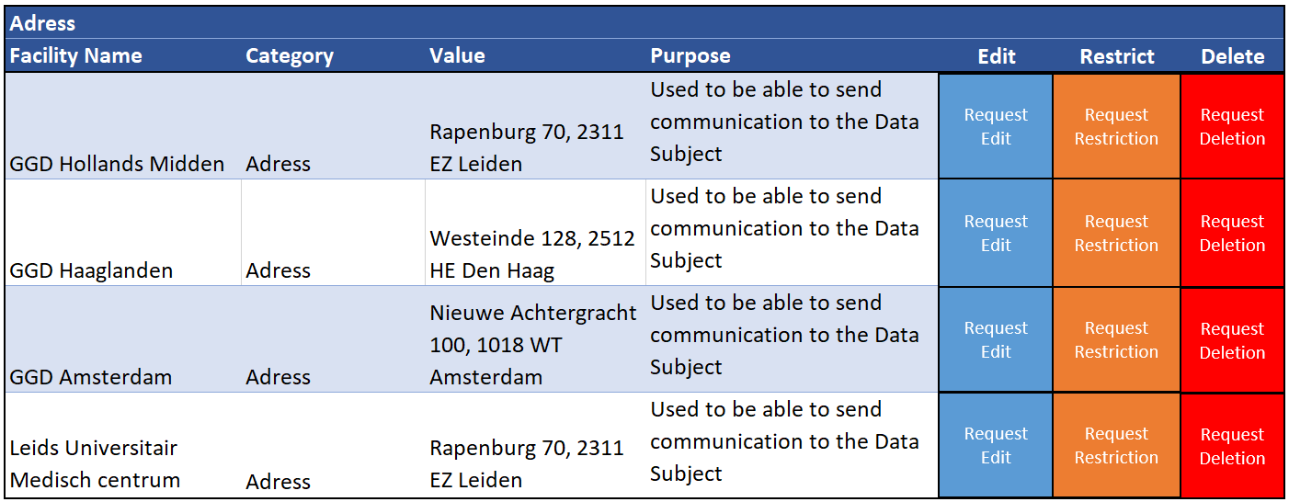
Since knowing that a person is present in a specific system can be sensitive information, it might not be secure enough to simply inquire if a BSN or other identifier exists in a data set. Using a DigID request for each system is not feasible, as this method is intended as an alternative when the previous method would be unfeasible. However, it might be possible to use a single DigID request at the central registry to verify the data subject's identity. With this verification, the registry could send requests to each system to determine if any of them contain information related to the data subject. As this approach only discloses if the information is located in a system rather than revealing the actual data, it might be sufficiently secure, provided that the system's security remains a top priority.

The last possible approach involves the central registry maintaining a record of which individuals belong to which data sets, circumventing the need to send requests to systems to determine the presence of their information. However, this introduces its own security vulnerabilities, as the list itself is sensitive information that requires protection, and creating a system that continuously updates a large data set with 18 million citizens belonging to multiple systems would be challenging.

8.2.4.2 Implementation

Assuming that the previous problem can be resolved, a solution of this kind would offer considerable value to the data subject. This section presents two possible approaches for creating a comprehensive overview of data usage across various organizations and systems. By providing this overview, the data subject is informed about the extent, value, and purpose of their data usage. Consequently, they can exercise their data ownership-related rights, which in this context include the right to request edits or rectifications for incorrect values, the right to restrict processing, and the right to request data deletion. Although various legal complexities prevent the system from being fully automated, necessitating some degree of organizational involvement, it represents a significant improvement over traditional methods such as contacting organizations by letter or email. Moreover, both approaches should incorporate a history of any interaction with data belonging to the data subject and the status of any GDPR requests that have been made.

The first approach, depicted in Figure 19, offers an overview organized by the variable in use. In this example, the data subject's address value is displayed across multiple organizations. From this example, it becomes apparent that the address value of the data subject is inconsistent among these organizations, likely hindering effective communication with the data subject. In cases like this, it is reasonable that organizations such as the GGD may not possess the correct address, as individuals typically belong to a single region unless they relocate. If the address is not updated upon relocation, the information will remain inaccurate. In theory, the basisregister system, shown in Figure 60, was designed to resolve such issues, but it has not been entirely successful with organizations still storing this information instead of linking to this central data source, resulting in a data duplication error. This approach is also unable to prevent such situations from occurring but does make it easier to identify inconsistencies when they occur, or even to notify the data subject when accessing their data via the central registry.

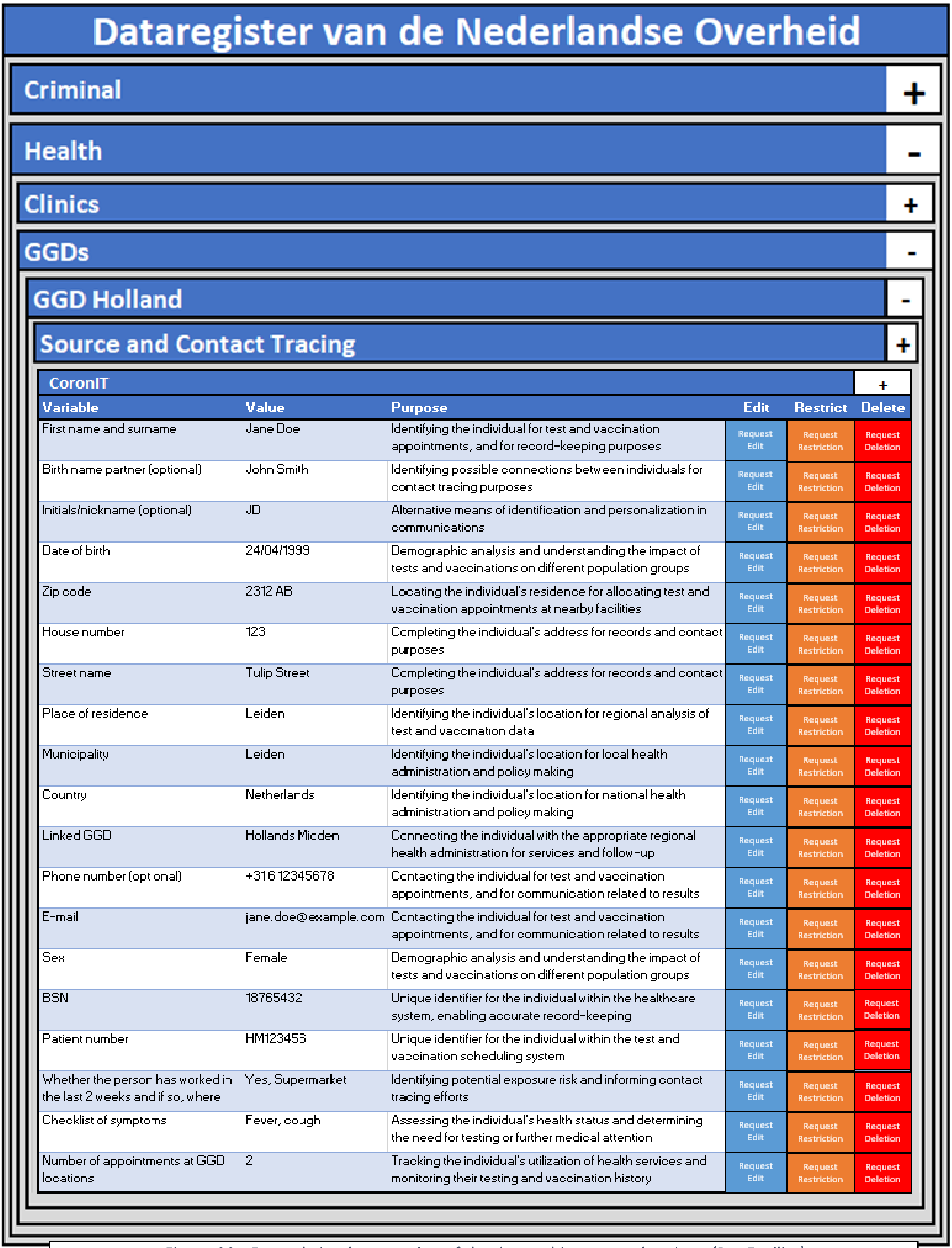


Adress						
Facility Name	Category	Value	Purpose	Edit	Restrict	Delete
GGD Hollands Midden	Adress	Rapenburg 70, 2311 EZ Leiden	Used to be able to send communication to the Data Subject	Request Edit	Request Restriction	Request Deletion
GGD Haaglanden	Adress	Westeinde 128, 2512 HE Den Haag	Used to be able to send communication to the Data Subject	Request Edit	Request Restriction	Request Deletion
GGD Amsterdam	Adress	Nieuwe Achtergracht 100, 1018 WT Amsterdam	Used to be able to send communication to the Data Subject	Request Edit	Request Restriction	Request Deletion
Leids Universitair Medisch centrum	Adress	Rapenburg 70, 2311 EZ Leiden	Used to be able to send communication to the Data Subject	Request Edit	Request Restriction	Request Deletion

Figure 19 - Example implementation of the data subject central registry (Per Variable)

The second approach, depicted in Figure 20, provides an overview organized by the system containing information about the data subject. This method offers a detailed view of all information related to the data subject within any specific system, meeting the requirements of the right of access by the data subject. In this example, the majority of values from CoronIT have been used to demonstrate how such a system would function.

The example could also be expanded significantly by including additional and/or more detailed information. For instance, the current example does not contain information about when or how often their data has been accessed, which is information that is already stored in the system to be able to identify suspicious behavior. It also does not contain information about how the variable is processed beyond its general purpose, which could be divided into multiple different use cases, which each return a different level of information to an employee.



Dataregister van de Nederlandse Overheid

Criminal +

Health -

Clinics +

GGDs -

GGD Holland -

Source and Contact Tracing +

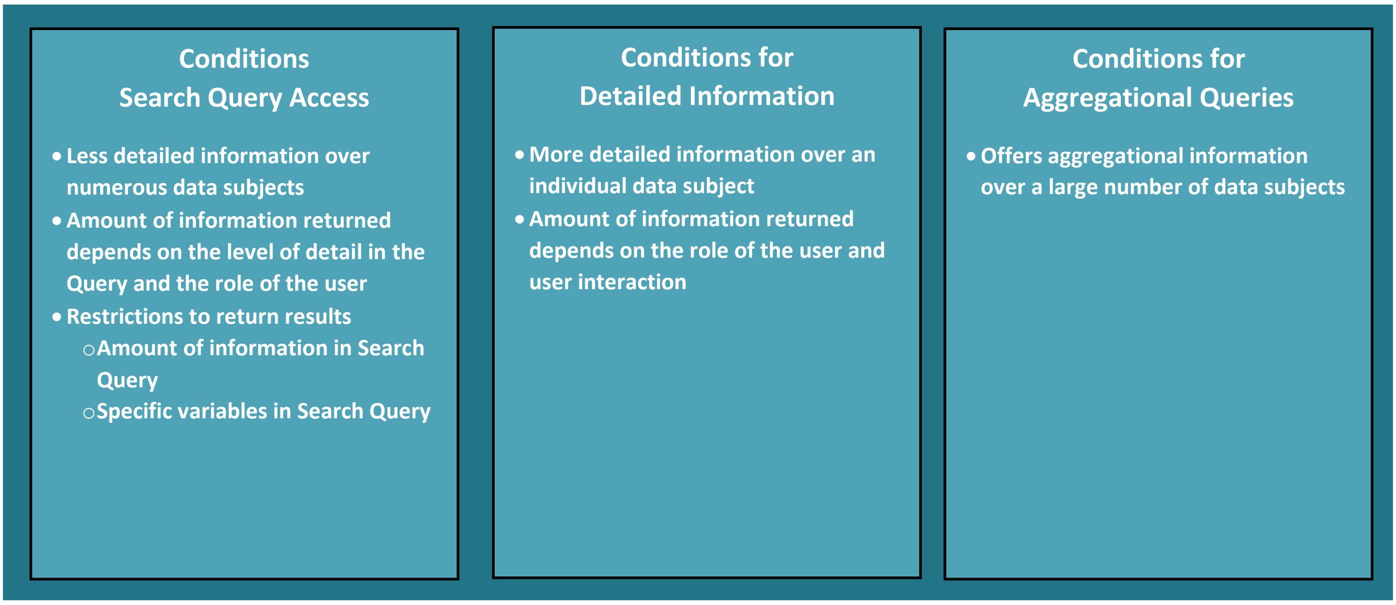
CoronIT +

Variable	Value	Purpose	Edit	Restrict	Delete
First name and surname	Jane Doe	Identifying the individual for test and vaccination appointments, and for record-keeping purposes	Request Edit	Request Restriction	Request Deletion
Birth name partner (optional)	John Smith	Identifying possible connections between individuals for contact tracing purposes	Request Edit	Request Restriction	Request Deletion
Initials/nickname (optional)	JD	Alternative means of identification and personalization in communications	Request Edit	Request Restriction	Request Deletion
Date of birth	24/04/1999	Demographic analysis and understanding the impact of tests and vaccinations on different population groups	Request Edit	Request Restriction	Request Deletion
Zip code	2312 AB	Locating the individual's residence for allocating test and vaccination appointments at nearby facilities	Request Edit	Request Restriction	Request Deletion
House number	123	Completing the individual's address for records and contact purposes	Request Edit	Request Restriction	Request Deletion
Street name	Tulip Street	Completing the individual's address for records and contact purposes	Request Edit	Request Restriction	Request Deletion
Place of residence	Leiden	Identifying the individual's location for regional analysis of test and vaccination data	Request Edit	Request Restriction	Request Deletion
Municipality	Leiden	Identifying the individual's location for local health administration and policy making	Request Edit	Request Restriction	Request Deletion
Country	Netherlands	Identifying the individual's location for national health administration and policy making	Request Edit	Request Restriction	Request Deletion
Linked GGD	Hollands Midden	Connecting the individual with the appropriate regional health administration for services and follow-up	Request Edit	Request Restriction	Request Deletion
Phone number (optional)	+316 12345678	Contacting the individual for test and vaccination appointments, and for communication related to results	Request Edit	Request Restriction	Request Deletion
E-mail	jane.doe@example.com	Contacting the individual for test and vaccination appointments, and for communication related to results	Request Edit	Request Restriction	Request Deletion
Sex	Female	Demographic analysis and understanding the impact of tests and vaccinations on different population groups	Request Edit	Request Restriction	Request Deletion
BSN	18765432	Unique identifier for the individual within the healthcare system, enabling accurate record-keeping	Request Edit	Request Restriction	Request Deletion
Patient number	HM123456	Unique identifier for the individual within the test and vaccination scheduling system	Request Edit	Request Restriction	Request Deletion
Whether the person has worked in the last 2 weeks and if so, where	Yes, Supermarket	Identifying potential exposure risk and informing contact tracing efforts	Request Edit	Request Restriction	Request Deletion
Checklist of symptoms	Fever, cough	Assessing the individual's health status and determining the need for testing or further medical attention	Request Edit	Request Restriction	Request Deletion
Number of appointments at GGD locations	2	Tracking the individual's utilization of health services and monitoring their testing and vaccination history	Request Edit	Request Restriction	Request Deletion

Figure 20 - Example implementation of the data subject central registry (Per Facility)

8.3 System Implementation to Limit Data Access to Information Need

Instead of returning the full data entry of a data variable by default, which would go against the principles of privacy and security by design, and data minimalization, data should instead be returned to the user based on their informational need. Section 8.1 described the content and the design of the to be created general Dutch governmental ontology according to which data is structured, Section 8.2 provided various use cases of this information in the context of a central registry of information. This section describes the way the information from the first section can expand upon to limit the extent of information revealed, in the context of processing the actual personal data in a facility. This section is divided into three parts, which correspond with three expected main use cases of any governmental healthcare system, as illustrated in Figure 21.



Conditions Search Query Access

- Less detailed information over numerous data subjects
- Amount of information returned depends on the level of detail in the Query and the role of the user
- Restrictions to return results
    - Amount of information in Search Query
    - Specific variables in Search Query

Conditions for Detailed Information

- More detailed information over an individual data subject
- Amount of information returned depends on the role of the user and user interaction

Conditions for Aggregational Queries

- Offers aggregational information over a large number of data subjects

Figure 21 - Components of the Metadata Description

Section 8.3.1 outlines the Accessibility of data used in the Search Query system, which identifies specific data subjects and enables further access through subsequent components. A metadata conditions table determines the fields accessible to different user roles, the requirements to initiate a search, and the extent of the information returned to the user. Additionally, this table can be used to automatically create the Search Query UI itself. This section includes a sample metadata conditions table, a flowchart, and pseudocode for the system.

Section 8.3.2 outlines the Accessibility of data when more detailed information is needed, such as accessing a patient dossier in the medical field. A metadata conditions table specifies whether a user role can access the relevant patient dossier and the amount of information disclosed in that request. The table also determines if user interaction is permitted for specific user roles to obtain more detailed information. This section includes a sample metadata conditions table, a flowchart, and pseudocode for both use cases.

Section 8.3.3 outlines the Accessibility of data used in the Search Query system for identifying trends, generating current information, updating dashboards, and facilitating deeper investigations by users. A metadata conditions table defines the fields accessible to different user roles, the level of the information returned, and a minimum threshold to ensure data anonymity. This section includes a sample metadata conditions table, a flowchart, and pseudocode for the system.

8.3.1 Search Query Access

This section discusses the implementation of a FAIR-based framework in a search query system for identifying patient records. The framework governs all user interactions with the system through metadata description tables associated with the data. The metadata description table in this section provides information on the variables available to users for creating queries, the data format returned to users, and the variable sensitivity.

Table 15 presents additional metadata for each variable in the database used for this search query system. The metadata includes a description of the variable's function in the context of the search query, its sensitivity (which depends on the variable itself and its use in the specific case), and the variable format that the user must follow. The metadata may differentiate based on the user's role, with different users having varying access levels and conditions. Section 8.5.1.1.offers a more detailed explanation of this user role system.

The conditions under which information may be returned are described in the table, with a practical implementation illustrated in Section 8.5.3. The implementation depends on the inclusion of specific fields and/or the number of fields as well as the number of results returned by the search query. The number of results is used to determine the level of detail of the information returned to the user. A data transformation system, as described in Section 8.5.4, is employed for this purpose.

	Value	Explanation
System Name	General_Search	Name of the System these conditions apply to
Type of Table	Search	Description of the type of system these conditions apply to
Variable Name	Variable X	The variable to which these conditions apply
Sensitivity	High	Sensitivity of the variable in this specific context
Available in Search?	Role X: No Role Y: Yes Role Z: Yes	A condition determining if the variable is accessible in the search Query, depending on the role of the user
Search Form	Role Y: Role Z:	A condition determining the form of the variable in the search query, depending on the role of the user. (YYYY-MM-DD, Integer for years or age, strings for surnames such as ‘Van der Plas’/’Plas, Van der’, last digits, per digit entree, etc.)
Must be included?	Role Y: Yes Role Z: No	A condition requiring this variable to be included in the search query before results may be returned, depending on the role of the user.
Minimal number of fields	Role Y: 4 Role Z: 4	A condition requiring a certain number of variables to be included in the search query before results may be returned, depending on the role of the user.
The threshold for Low information	Role Y: 50 Role Z: 100	A threshold that determines when a low amount of information can be revealed, depending on the role of the user
Low information	Role Y: XX Role Z: YY	The form of the information returned to the user when a low amount of information can be revealed, depending on the role of the user
The threshold for moderate information	Role Y: 20 Role Z: 50	A threshold that determines when a moderate amount of information can be revealed, depending on the role of the user
Moderate Information	Role Y: XX Role Z: YY	The form of the information returned to the user when a moderate amount of information can be revealed, depending on the role of the user
The threshold for Detailed information	Role Y: 10 Role Z: 20	A threshold that determines when detailed information can be revealed, depending on the role of the user
Detailed Information	Role Y: XX Role Z: YY	The form of the information returned to the user when detailed information can be revealed, depending on the role of the user

Table 15 - Metadata Conditions for the Search System

Figure 22 presents a flowchart of the Search Query system, which consists of four components: the user, the security validator system, the metadata table, and the database. The process begins with the user initiating a request by sending a query containing their role, credentials, and intended search query information. The security validator verifies the user's role and credentials. If incorrect, an error is returned, and the monitoring system takes further action. If the credentials are correct, the process proceeds to the metadata table.

The metadata table component evaluates the query to determine if it meets the required conditions for initiating a search and whether the search query contains the necessary information. If the query includes information that the user's role does not have access to, the query is either deemed incorrect and an error is returned or the information is disregarded. If the query meets the required conditions, it is sent to the database, where initially only the number of results is returned.

The number of results is compared to various thresholds listed in the metadata table. If the lowest threshold is not exceeded, an error is returned to the user. If the threshold is exceeded, another request is sent to the database, returning the actual results but limited by the constraints listed in the metadata table. This response is then returned to the user, who can either refine their search query or access specific patient dossiers.

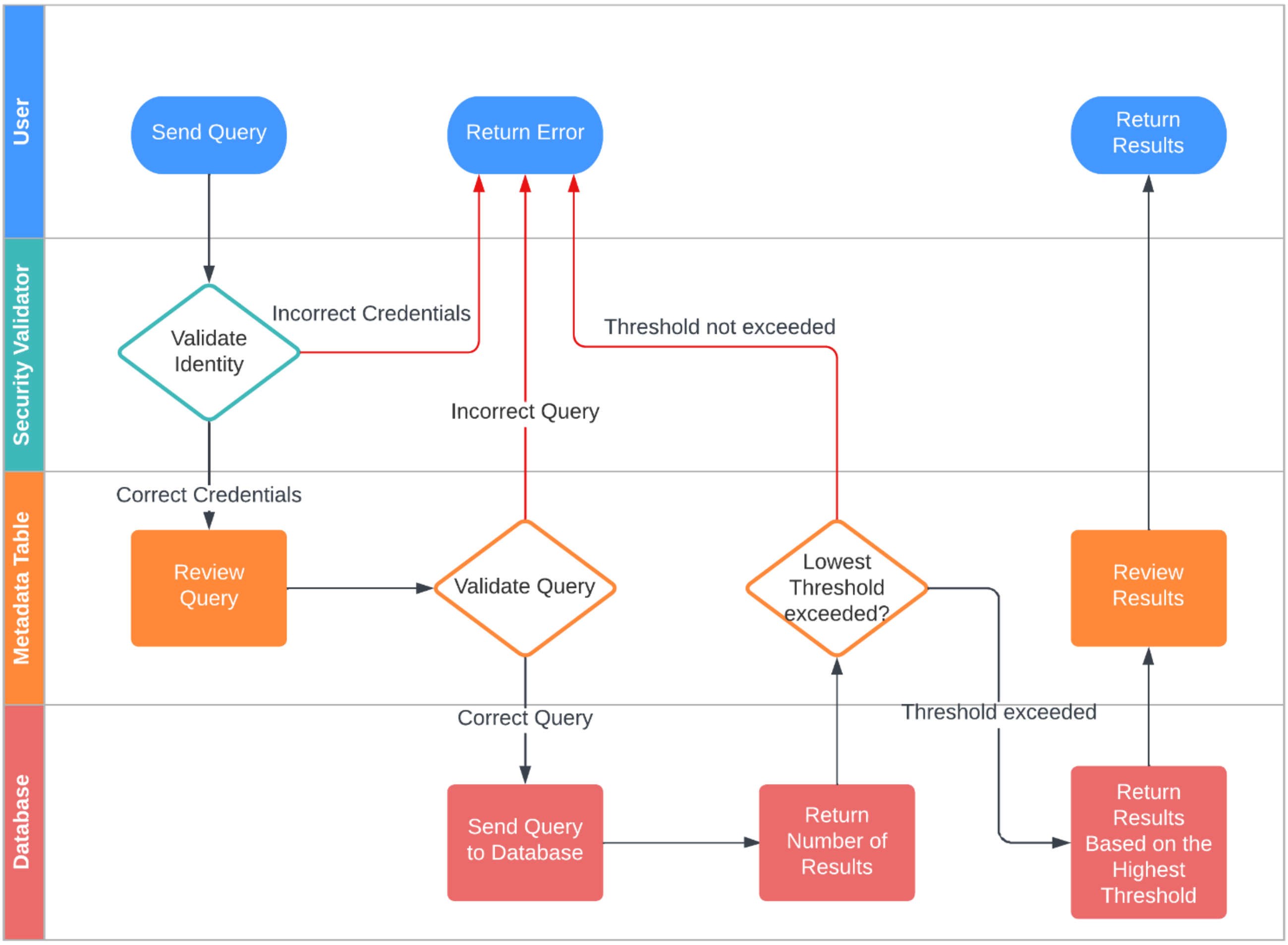


Figure 22 - Flowchart of the Search System

The proposed system ensures that information is returned to the user in a privacy- and security-focused manner. If search systems are required for different system sections, another metadata description table can be created using a different API to describe the conditions for various use cases while adhering to the same principles.

The metadata conditions table can also serve as a tool for automatically creating the search system UI and API based on the processing requirements of a specific use case. This approach ensures that the metadata conditions table accurately reflects the organization's data processing needs and enables quick changes to be made without requiring technical expertise.

Within the search system UI context, three variables are essential: "Variable Name," "Available in Search?", and "Search Form." The "Variable Name" corresponds to the respective variable and acts as the label for the data entry field. The "Available in Search?" variable determines the accessibility of variables to different user roles, ensuring data security. The "Search Form" variable indicates the proper data entry format for each variable. For instance, a data subject's surname would be input in a simple string field, while various formats could be used to enter a date of birth, such as selecting the month, day, and year, or using a simplified format like DD-MM-YY or the data subject's age in years. Similarly, a data subject's BSN could be entered in various formats, such as eight 1-digit entry fields or a 3-digit entry field for the BSN's last three digits.

The "Must be included?" and "Minimal number of fields" variables are primarily used in the API, which will be discussed later in the section. Nevertheless, these variables also offer crucial information to the user regarding the search system requirements. Section 8.5.3.1 provides a more detailed explanation of this implementation.

The API's pseudocode for the search system is illustrated in Figure 23, aligning with the flowchart in Figure 22. The pseudocode depends on the creation of other functions; however, these functions are independent of the actual data being processed. These functions can be continuously developed, independent of any specific project, to further enhance the security of the system.

```
def General_Search_Search(Identifier, Role, User_Query):
    Metadata_Conditions_Table = Tables_data(General_Search)

    #This function validates if the identifier of the user is valid and if 
their role matches this identifier
    if not Validate_identity(Identifier, Role)
        return("Error: Invalid Credentials")

    #This function validates the query of the user, based on the format and 
if the user role has access to these specific fields as defined in the 
metadata conditions table
    if not Validate_Query(User_Query, Role, Metadata_Conditions_Table)
        return("Error: Invalid Query due to X")

    #This function converts the query of the user into a query that only 
returns the number of results before storing this number
    Database_Query = Query_Formatter_Number_of_Results(User_Query)
    Number_of_results = Database_Function(Database_Query)

    #This function determines if the number of results returned by the 
query meets the thresholds set to return information.
    if not Threshold(Number_of_results, Role, Metadata_Conditions_Table):
        return("Error: Threshold not exceeded, Query is too generic")
    #If the threshold set to return information is met, the system 
determines the level of information that can be returned based on the 
number of results and the user role
    else:
        level_of_information = Threshold_level(Number_of_results, Role, 
Metadata_Conditions_Table)

    #This function converts the query of the user into a query that returns 
results, restricted by the level of information that can be exposed
    Database_Query = Query_Formatter_Level_of_Information(User_Query, 
level_of_information)

    #Another Query is sent to the database, this time restricted by the 
level of information that can be exposed
    Results = Database_Function(Database_Query)

    return(Results)
```

Figure 23 - Pseudocode for the Search System

.

8.3.2 Detailed information – Patient Dossier

This section describes the implementation of the FAIR-based framework in a patient dossier system used to provide detailed information about a specific data subject. The framework regulates all user interactions with the system through metadata description tables attached to the data. There are two types of metadata description tables used in this section, with the first being used to regulate the extent to which information is provided to the user, and the second being used to provide the user with access to functionalities and limit the amount of information exposed through this process.

To increase security even further, accessing a patient dossier could be limited to only being accessible via the search query system. This could be achieved by making use of the information from the search query results and adding this information to the query to access the patient dossier as a form of validation. As a patient dossier would normally be accessed only after searching for a specific user, the impact this would have on users is minimal. However, malicious actors would be unable to make use of the API to access patients' dossiers as they would lack the required validation information. Through this, the malicious actor would be forced to make use of the search system, thereby allowing the monitoring system to spot their behavior.

8.3.2.1 Accessing Information

Table 16 presents additional metadata for each variable in the database used for returning information in the patient dossier system. The metadata primarily dictates the level of information that is provided to the user without user interaction. The metadata may also make a distinction based on the role of the user, with different users having different levels of access, and with different conditions. A more detailed explanation of this system of user roles can be found in Section 8.5.1.1.

	Value	Explanation
System Name	Access Patient Dossier	Name of the System these conditions apply to
Type of Table	Detailed Information	Description of the type of system these conditions apply to
Variable	Variable X	The variable to which these conditions apply
Sensitivity	High	Sensitivity of the variable in this specific context
Available in overview?	Role X: No Role Y: Yes Role Z: Yes	A condition determining if the variable is included in the overview, depending on the role of the user
Level of information before user interaction	Role Y: Yes Role Z: Yes	The form of the information returned to the user when a patient dossier is accessed, depending on the role of the user

Table 16 - Metadata Conditions for the Patient Dossier

A flowchart of the initial opening of the patient dossier is presented in Figure 24. The system consists of four lanes, namely the user, the security validator system, the metadata table, and the database. The process starts with the user initiating a request by interacting with the functionality to open a patient dossier in the Search Query menu, which includes their role, credentials, and the exposed information belonging to that data subject. The security validator validates the role and credentials of the user. If they are incorrect, an error is returned, and the monitoring system takes further action. If the credentials are correct, the process continues to the metadata table.

The metadata table component verifies the request by testing if the exposed information from the Search Query menu matches the information from the patient dossier the user is attempting to open. If this information doesn't match then the request is invalid and the monitoring system takes further action. This is especially important here as this check should only be invalidated if the user was attempting to bypass the search query system.

If the information does match the information from the patient dossier, then a final check is done to verify that the user can access the information from the specific patient record, which is reliant on the role of the user although additional patient records specific flags could be made in certain instances. If the role of the user is unable to access the patient dossier, then an error is returned, else the system continues to the final step.

The final step is to send another request to the database, returning the actual results but limited by the constraints listed in the metadata table. This response is then returned to the user, who can use the information presented to him for their responsibilities, or make further use of this system to access functionalities or request for more information to be made available.

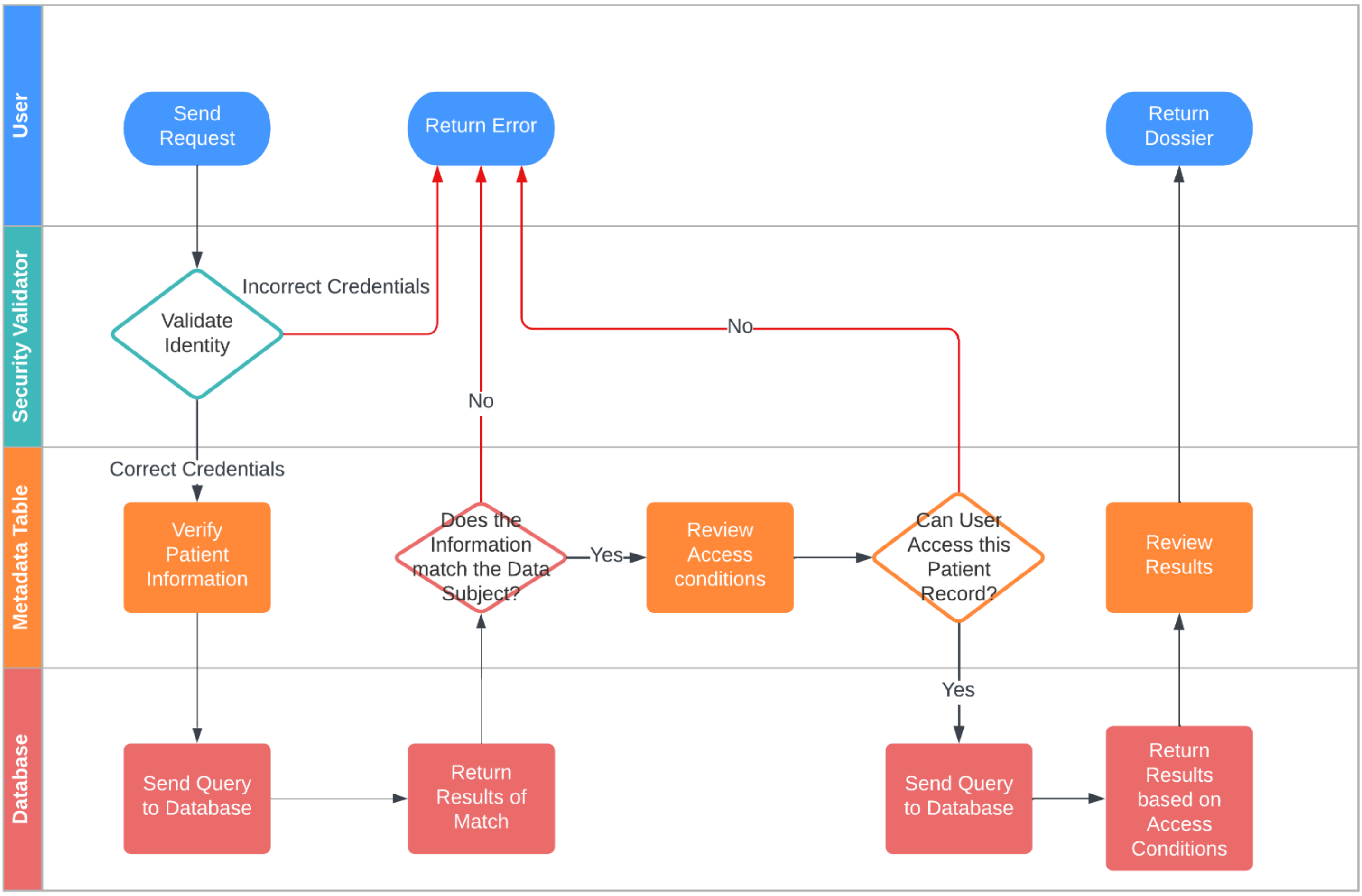


Figure 24 - Flowchart of the Initial Opening of the Patient Dossier

The metadata conditions table can serve as a tool to create the API to retrieve the Patient Dossier while limiting the amount of information that is exposed to the user. The Pseudocode of this approach has been depicted in Figure 25, with the only difference between different projects and different data being contained within the metadata conditions table.

```
def Patient_Dossier_Access(Identifier, Role, User_Query):
    Metadata_Conditions_Table = Tables_data(Access_Patient_Dossier)

    #This function validates if the identifier of the user is valid and if their role matches this identifier
    if not Validate_identity(Identifier, Role)
        return("Error: Invalid Credentials")

    #This function validates if the information in the request matches the information contained in the database
    if not Validate_Information(Patient_Dossier_Key, Patient_Information)
        return("Error: Invalid request, Patient Information does not Match Patient Dossier")

    #This function validates if the user role has access to the patient record
    if not Validate_Access(Role, Metadata_Conditions_Table)
        return("Error: Invalid request, the user role does not have access to this patient dossier")

    #This function retrieves the level of information the user role has access to
    level_of_information = Threshold_level(Role, Metadata_Conditions_Table)

    #This function generates the query sent to the database to retrieve information from the patient dossier, restricted by the level of information that can be exposed
    Database_Query = Query_Formatter_Retrieve_Patient_Dossier(level_of_information)

    #Another Query is sent to the database, this time restricted by the level of information that can be exposed
    Patient_Dossier = Database_Function(Database_Query)

    return(Patient_Dossier)
```

Figure 25 – Pseudocode for Opening a Patient Dossier

The same approach used to increase security when accessing a patient dossier may also be applied to the system that allows users to receive more information upon user interaction. Given that this system should only be used from the patient dossier menu, the information in the patient dossier can be used as a validator to receive more detailed information.

Table 16 presents additional metadata for each variable in the database used for returning information in the patient dossier system. The metadata primarily dictates the level of information that is provided to the user without user interaction, if user interaction is possible, what level of information is provided to the user upon this user interaction, and if the value can be adjusted by the user. The metadata may also make a distinction based on the role of the user, with different users having different levels of access, and with different conditions. A more detailed explanation of this system of user roles can be found in Section 8.5.1.1.

	Value	Explanation
System Name	Patient Dossier Access Detailed	Name of the System these conditions apply to
Type of Table	Detailed Information	Description of the type of system these conditions apply to
Variable	Variable X	The variable to which these conditions apply
Sensitivity	High	Sensitivity of the variable in this specific context
Is user interaction possible?	Role Y: Yes Role Z: Yes	A condition determining if the variable can expose more information upon user interaction, depending on the role of the user
Level of information after user interaction	Role Y: XX Role Z: YY	The form of the information returned to the user when the variable is exposed through user interaction, depending on the role of the user
Are user adjustments possible?	Role Y: No Role Z: Yes	A condition determining if the variable can be adjusted by the user, depending on the role of the user

Table 17 - Metadata Conditions for user interaction in the Patient Dossier

A flowchart of user interaction used to increase the level of information in the patient dossier is presented in Figure 26. The system consists of four lanes, namely the user, the security validator system, the metadata table, and the database. The process starts with the user initiating a request by interacting with the functionality that allows for more information to be exposed, which includes their role, credentials, and the exposed information belonging to that data subject. The security validator validates the role and credentials of the user. If they are incorrect, an error is returned, and the monitoring system takes further action. If the credentials are correct, the process continues to the metadata table.

The metadata table component verifies the request by testing if the exposed limited amount of information from the patient dossier matches the more detailed information from the patient dossier the user is attempting to request. If this information doesn't match then the request is invalid and the monitoring system takes further action. This is especially important here as this check should only be invalidated if the user was attempting to bypass the system used to request more detailed information.

If the information does match the information from the patient dossier, then a final check is done to verify that the user can access more detailed information and in what form, which is reliant on the role of the user. If the role of the user is unable to access more detailed patient information, then an error is returned, else the system continues to the final step.

The final step is to send another request to the database, returning the more detailed information in the form of the format listed in the metadata table. This response is then returned to the user, who can use the information presented to him for their responsibilities.

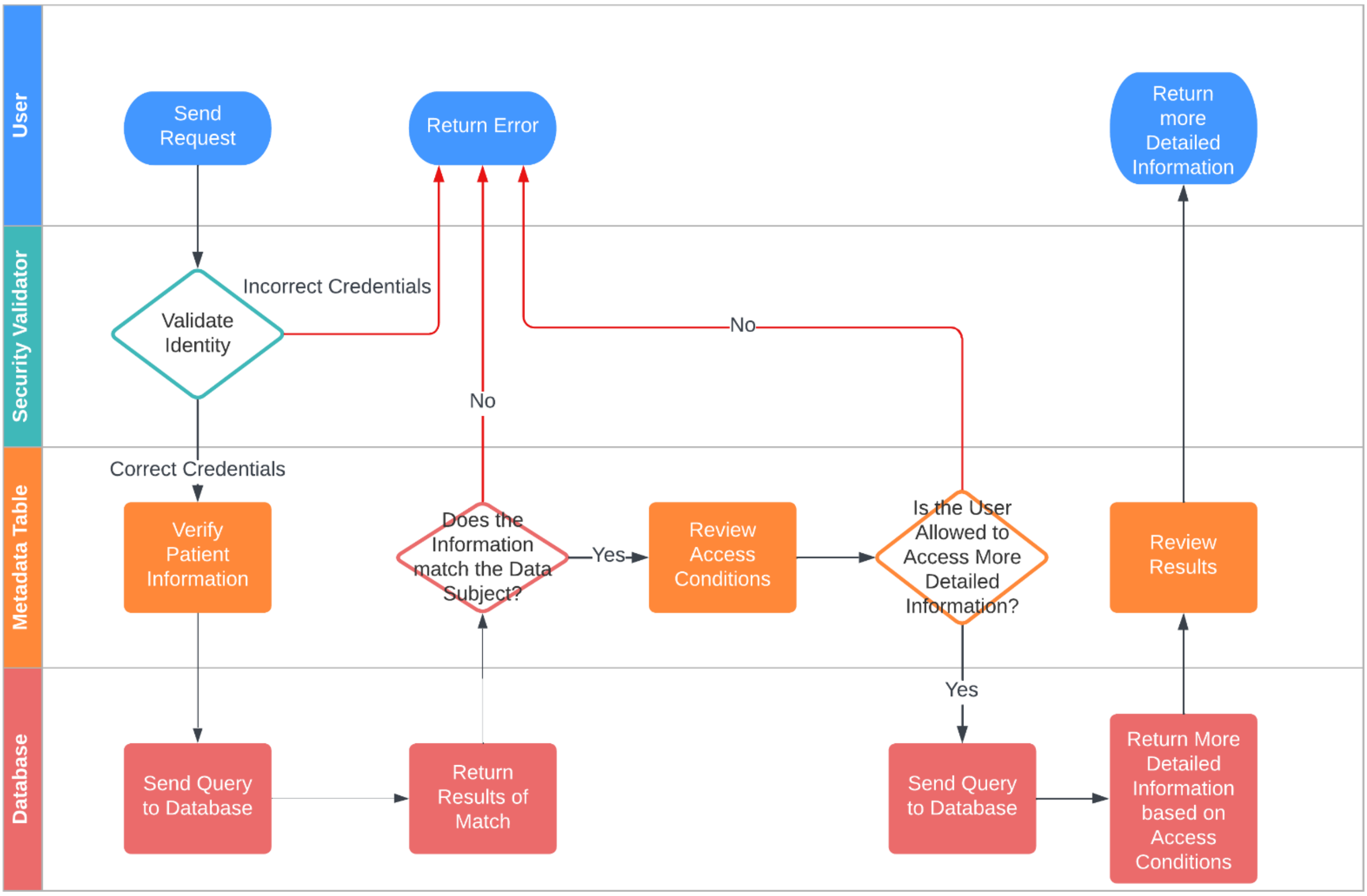


Figure 26 - Flowchart of the user interaction used to receive more detailed information in the Patient Dossier

The pseudocode described in Figure 27 for disclosing additional patient information bears a strong resemblance to the pseudocode used to gain entry into the patient dossier, as shown in Figure 25. The sole differentiation lies in the fact that instead of authenticating the user's eligibility to access the patient dossier, a corresponding check is performed to verify whether the user can access more detailed information than is currently available. Upon successful verification, the same processes as in Figure 25 are executed, but with an elevated access level. Resulting in the user being provided with more detailed information about the data subject.

```
def Patient_Dossier_Access_Detailed(Identifier, Role, User_Query):
    Metadata_Conditions_Table = Tables_data(Access_Patient_Dossier)

    #This function validates if the identifier of the user is valid and if their role matches this identifier
    if not Validate_identity(Identifier, Role)
        return("Error: Invalid Credentials")

    #This function validates if the information in the request matches the information contained in the database
    if not Validate_Information(Patient_Dossier_Key, Patient_Information)
        return("Error: Invalid request, Patient Information does not Match Patient Dossier")

    #This function validates if the user role has access to more detailed information
    if not Validate_Additional_Access(Role, Metadata_Conditions_Table)
        return("Error: Invalid request, the user role does not have access to more detailed information")

    #This function retrieves the level of information the user role has access to
    level_of_information = Threshold_level(Role, Metadata_Conditions_Table, Detailed=True)

    #This function generates the query sent to the database to retrieve information from the patient dossier, restricted by the level of information that can be exposed
    Database_Query = Query_Formatter_Retrieve_Patient_Dossier(level_of_information)

    #Another Query is sent to the database, this time restricted by the level of information that can be exposed
    Patient_Dossier = Database_Function(Database_Query)

    return(Patient_Dossier)
```

Figure 27 – Pseudocode for exposing additional information in the Patient Dossier

8.3.2.2 Using functions

Table 18 presents additional metadata for each variable in the database able to be used for any specific function in the patient dossier system. Every function would have its own metadata table, specifying for each variable what the sensitivity of the variable is in the current context if the variable is allowed to be used by the user in the specific function, and what level of information would be provided to the user within that function. The metadata may also make a distinction based on the role of the user, with different users having different levels of access, and with different conditions. A more detailed explanation of this system of user roles can be found in Section 8.5.1.1.

	Value	Explanation
System Name	Patient Dossier Function X	Name of the System these conditions apply to
Type of Table	Function	Description of the type of system these conditions apply to
Variable	Variable X	The variable to which these conditions apply
Sensitivity	High	Sensitivity of the variable in this specific context
Available in function?	Role X: No Role Y: Yes Role Z: Yes	A condition determining if the variable can be used in the function, depending on the role of the user. This dictates if the user can make use of the function in the first place.
Level of information in function?	Role Y: XX Role Z: YY	The form of information returned to the user when the function is used, depending on the role of the user

Table 18 - Metadata Conditions for any function in the Patient Dossier

The same approach used to increase security in the rest of the patient dossier may also be applied to functions related to the patient dossier. Given that this system should only be used from the patient dossier menu, the information in the patient dossier can be used as a validator to ensure that the correct workflow is followed.

A flowchart of the use of functions in the patient dossier is presented in Figure 28. The system consists of four lanes, namely the user, the security validator system, the metadata table, and the database. The process starts with the user initiating a request by interacting with the functionality used to access any specific function menu. The security validator validates the role and credentials of the user. If they are incorrect, an error is returned, and the monitoring system takes further action. If the credentials are correct, the process continues to the metadata table.

The metadata table component verifies the request by testing if the exposed limited amount of information from the patient dossier matches the information from the patient dossier. If this information doesn’t match then the request is invalid and the monitoring system takes further action. This is especially important here as this check should only be invalidated if the user was attempting to bypass the system used to write information to the database itself.

If the information does match the information from the patient dossier, a check is done to verify that the user has access to the functionality, which is reliant on the role of the user. If the role of the user is unable to access the function, the request is denied and an error is returned to the user.

If the user can access the function, another check is done to determine the amount of information the function menu is allowed to expose. If the user is allowed access to information, a request is sent to the database, retrieving the amount of information that can be exposed to the user. Based on this, the user is always presented with a function menu, although the level of information can differ between roles and may also not expose any information. From this menu, the user can make use of the functionality and send a statement to the database, either creating new information or updating existing information.

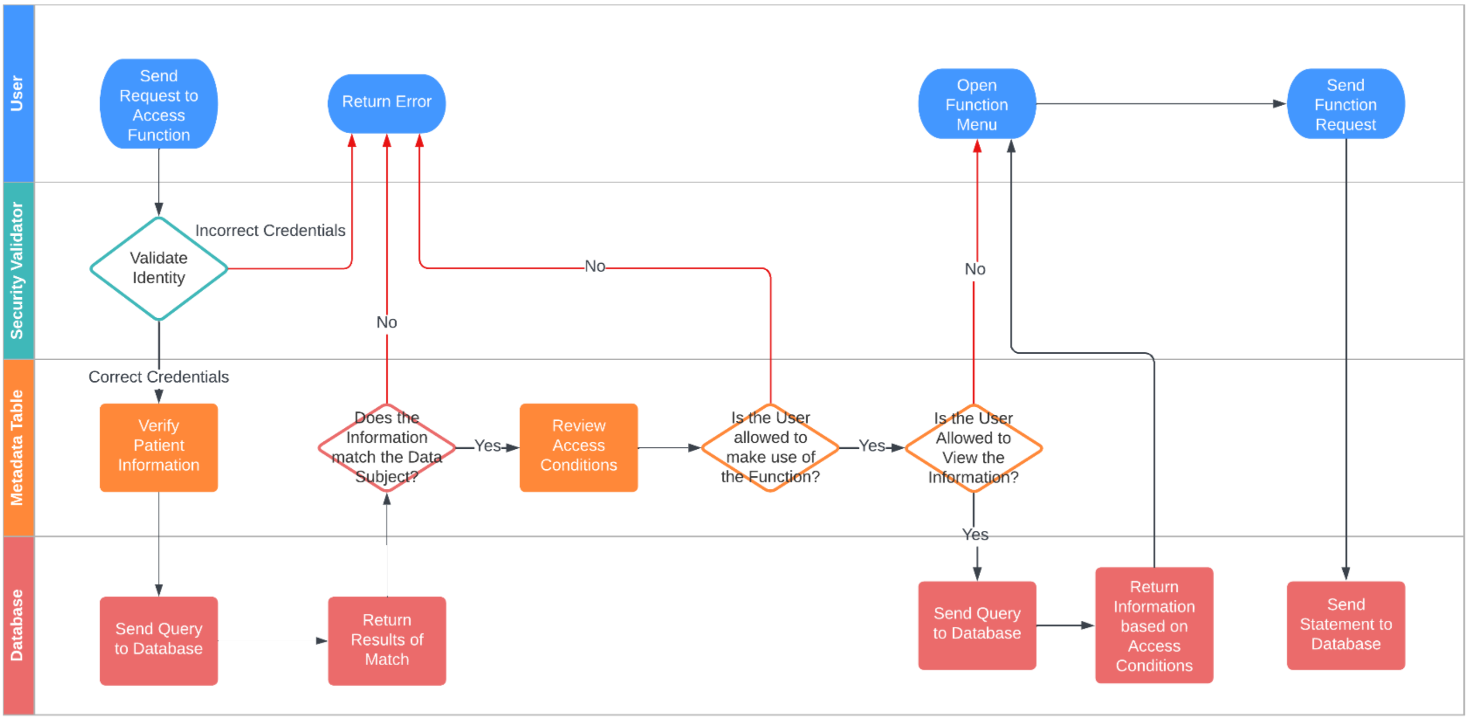

User
Security Validator
Metadata Table
Database
Send Request to Access Function
Return Error
Open Function Menu
Send Function Request
Validate Identity
Incorrect Credentials
No
No
No
Correct Credentials
Verify Patient Information
Does the Information match the Data Subject?
Yes
Review Access Conditions
Is the User allowed to make use of the Function?
Yes
Is the User Allowed to View the Information?
Yes
Send Query to Database
Return Results of Match
Send Query to Database
Return Information based on Access Conditions
Send Statement to Database

Figure 28 - Flowchart of the use of functions in the Patient Dossier

The metadata conditions table can serve as a tool to generate APIs for diverse functions, while also controlling the amount of information that is visible to the user. However, this process requires a distinct approach compared to previous elements because functions are specific to each project and cannot be entirely automated. To address this, the FAIR-based Framework standardizes aspects unrelated to the function itself, while providing developers with the ability to define how data should be utilized and modified only in certain sections. This allows developers to create any function needed for the program while adhering to the metadata conditions table requirements. Specifically, developers are unable to generate their own database calls or transactions and instead must use the information from preceding lines and submit the query to the validator that is automatically generated at the end of the function.

The Pseudocode of this approach has been depicted in Figure 25, which includes two distinct functions. The initial function controls the display of the function menu and the information that is presented to the user. The second function includes the developer's personal code, which is finalized by a validator to ascertain if the user can send the database transaction, and then returns the transaction outcome to the user. This aligns with the flowchart displayed in Figure 28.

```
def Patient_Dossier_Function_X_Open_Function_Menu(Identifier, Role, Patient_Dossier_Key, Patient_Information):
    Metadata_Conditions_Table = Tables_data(Patient_Dossier_Function_X)
    #This function validates if the identifier of the user is valid and if their role matches this identifier
    if not Validate_identity(Identifier, Role)
        return("Error: Invalid Credentials")

    #This function validates if the information in the request matches the information contained in the database
    if not Validate_Information(Patient_Dossier_Key, Patient_Information)
        return("Error: Invalid request, Patient Information does not Match Patient Dossier")

    #This function validates if the user role has access to the function
    if not Validate_Access(Role, Metadata_Conditions_Table)
        return("Error: Invalid request, the user role does not have access to this function")

    #This function validates if the user role has access to detailed information, before creating a menu instance
    if not Validate_Detailed_Information(Role, Metadata_Conditions_Table):
        Menu = Function_Menu(Detailed_Information=False)
    else:
        Menu = Function_Menu(Detailed_Information=True)

    return(Menu)

def Patien_Dossier_Function_X_Function_Menu(Previous_Information, Role):
    Metadata_Conditions_Table = Tables_data(Patient_Dossier_Function_X)

    # Write own code for functionalities, which is unable to execute any database transactions
    # The developer only has access to the data already accessed in the previous automatically generated lines
    # Database_Query = X

    #This function determines if the user is able to execute the specific transaction
    if not Database_Access(Database_Query, role, Metadata_Conditions_Table)
        return("Error: User Role is not allowed to execute this transaction")

    #The Database Transcation is sent to the database, if this fails an error is returned
    try:
        Database_Function(Database_Query)
    except:
        return("Error: Transaction was unsuccesfull")

    return(Results)
```

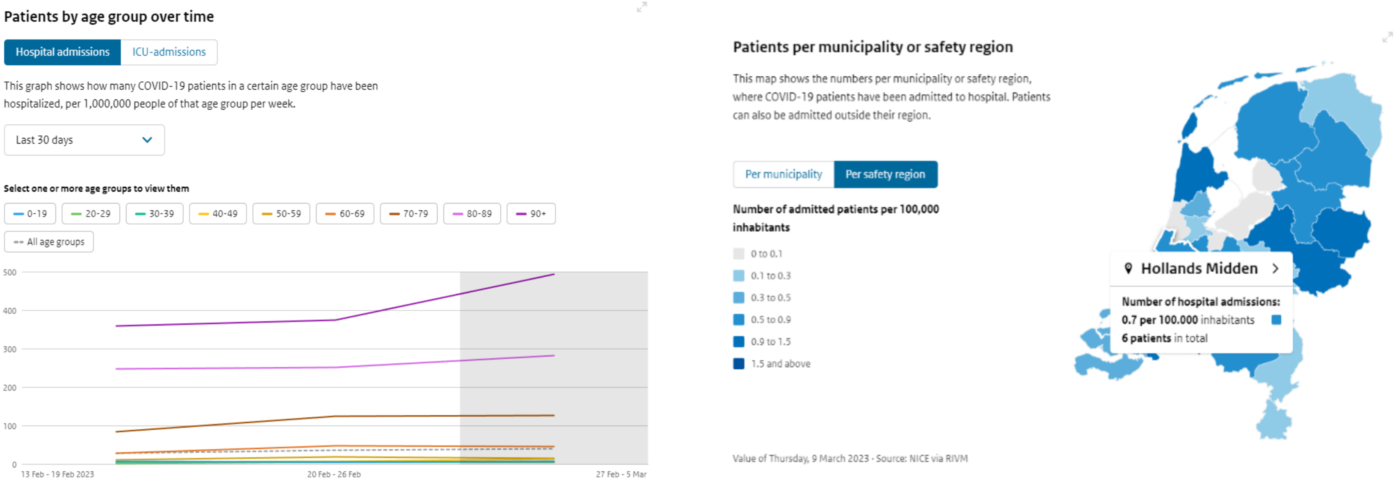
Figure 29 - Pseudocode for Functions in the Patient Dossier

8.3.3 Aggregational Queries

This section describes the implementation of the FAIR-based framework in the context of aggregational queries, which are utilized by organizations to obtain information regarding trends and specific groups. The framework serves as a regulatory mechanism for all user interactions with the system by employing a metadata description table that is linked to the data. Within this particular system, the metadata description table is constrained in its scope, given that the extraction of statistical data from the data through data visitation is already an inherent component of the FAIR framework. Consequently, the metadata description table is solely utilized in instances where data visitation is incapable of ensuring data security.

Depending on the nature of the data contained in the system and its relevance to the public, scheduled queries could be created that, through the principle of data visitation, provide either the public or public institutions with valuable information without them having ever received access to the data. Thereby guaranteeing its security. However, in many of these instances, data has already been anonymized to such a degree as to remove the need for security altogether. This would make this an alternative solution to the creation of dashboards instead of being able to provide dashboards with new information.

For example, In the context of the Covid-19 pandemic, the Dutch government created Coronadashboard, which utilizes data from the RIVM, the LCPS, the CBS, the various GGDs, and Dutch hospitals [227]. Offering overviews of trends (Figure 30 - Left) [228], current information (Figure 30 - Right) [228], and in some instances the anonymized raw data files used to create the figures [229]. In this instance, the FAIR-based framework would be able to be used to generate the information used to generate this dashboard, but it would not offer any improvement as there is no privacy or security problem to improve in the first place. Organizations already use data visitation as there is no need to send anything but the results of any query to the dashboard.



Patients by age group over time

This graph shows how many COVID-19 patients in a certain age group have been hospitalized, per 1,000,000 people of that age group per week.

Select one or more age groups to view them

Patients per municipality or safety region

This map shows the numbers per municipality or safety region, where COVID-19 patients have been admitted to hospital. Patients can also be admitted outside their region.

Number of admitted patients per 100,000 inhabitants

Number of hospital admissions:
0.7 per 100.000 inhabitants
6 patients in total

Value of Thursday, 9 March 2023 · Source: NICE via RIVM

Figure 30 - Dashboard Example of a Trend (Left) and current information (Right)

The FAIR-based framework in the context of aggregational queries would be fully compatible with the previously described system, with the possibility of generating these kinds of dashboards automatically. The addition of metadata to FAIRify the data according to Section 0, would increase the usability of these kinds of anonymized raw data files, which may have significant value in some instances but is unlikely to be able to offer significant value in the limited amount of information used to create such dashboard overviews.

Instead, the main value of the FAIR-based framework is that it allows for the creation of such trends and current information overviews based on data that cannot be anonymized. With queries being able to be applied to limited groups using sensitive information that would normally not be allowed to be used. As not even employees at any organization would be required to see the data, privacy, and security cannot be compromised.

Table 19 presents additional metadata for each variable in the database able to be used for the aggregational query system. The metadata table specifies for each variable what the sensitivity of the variable is in the current context if the variable is allowed to be used by the user in the query if the variable is returned in the results, and what the minimum number of entrees are for results to be provided. The metadata may also make a distinction based on the role of the user, with different users having different levels of access, and with different conditions. A more detailed explanation of this system of user roles can be found in Section 8.5.1.1.

	Value	Explanation
Variable	Variable X	The variable to which these conditions apply
Sensitivity	Low	Sensitivity of the variable in this specific context
Available in Filter?	Role X: No Role Y: Yes Role Z: Yes	A condition determining if the variable can be used in the aggregational query filter, depending on the role of the user.
Access to specific values?	Role Y: Yes Role Z: Yes	A condition determining if specific values can be used in the aggregational query filter, depending on the role of the user. E.g., age between 22 and 25 instead of age > risk factor
Available in Results?	Role X: No Role Y: Yes Role Z: Yes	A condition determining if the variable can be included in the results of the aggregational query filter, depending on the role of the user.
Minimal number of entrees	Role Y: 100 Role Z: 50	A condition determining the minimal number of filter entrees there need to be before statistical results may be returned, depending on the role of the user. This prevents filters from becoming too specific, which would eliminate the anonymity aggregational statistical data offers.

Table 19 - Metadata Conditions for the aggregational query system

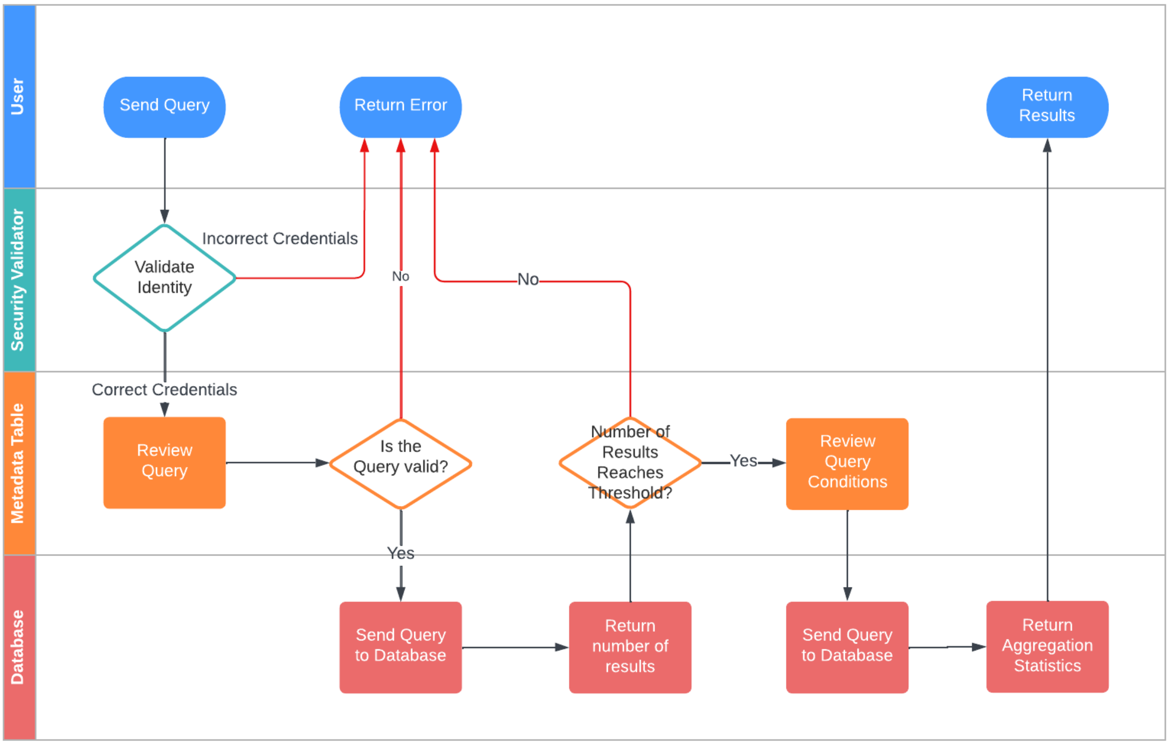


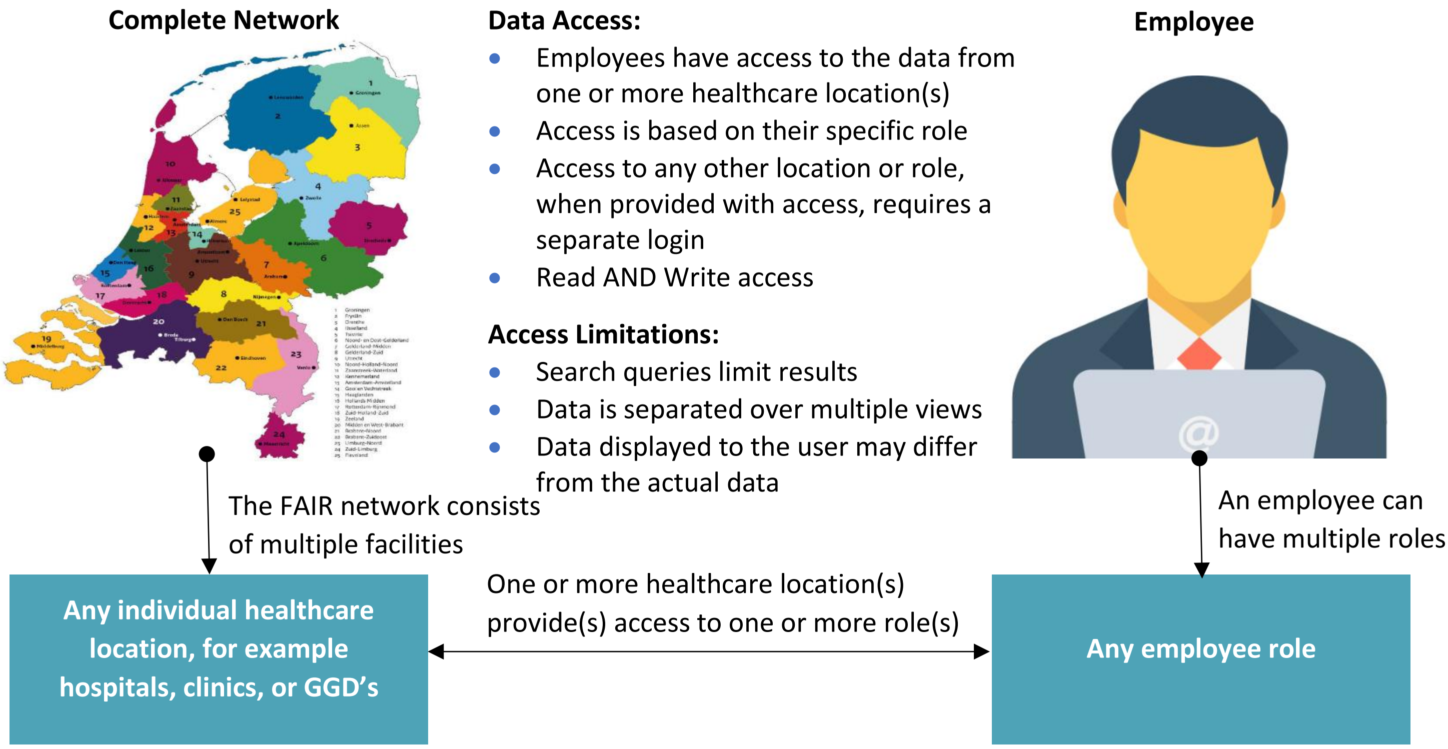
Figure 31 - Flowchart of the Aggregational Query

8.4 Data/Information Flow

This section will describe the data flow of a FAIR-based framework as it is being used in support of a healthcare activity. As such, it will involve the processing of personal data that has not been anonymized and can pose a significant risk for the data subject if the security or privacy of it was ever compromised. In contrast to processing data for scientific research, any such system will require both read and write access depending on the role of the employee that will be processing the data.

Given these conditions, the core concepts of FAIR alone aren't able to ensure that data is safely processed regarding the risk to the data subject. For example, a data visiting approach eliminates the need to share data files, which significantly increases security due to maintaining complete control over access to the data. The common standards allow for data to be used to support processes in many locations while ensuring that data has the same meaning everywhere. The clear agreements over the use of data ensure that data is only processed under approved conditions. But none of these concepts address any kind of internal risks from employees using their level of access to gain access to data beyond the confines of their responsibilities. At best, these FAIR concepts make it more difficult as there would not be any export functionality by default or monitoring can be significantly improved as organizations can control and monitor all access.

However, by extending the concepts of FAIR, you could minimize both the amount of information employees are exposed to and you could minimize the impact on a data subject if a personal data breach were to ever occur. In combination with the previous concepts, such an approach may even be able to able to completely prevent any malicious actor from even trying to leak information outside of the confines of the system as the risks of getting caught are increased and the gain from leaking information is significantly decreased. An Entity-Relationship Diagram of this has been depicted in Figure 32, with an in-detail explanation of various elements being described in Section 8.5.1.



Complete Network

Data Access:

- Employees have access to the data from one or more healthcare location(s)
- Access is based on their specific role
- Access to any other location or role, when provided with access, requires a separate login
- Read AND Write access

Access Limitations:

- Search queries limit results
- Data is separated over multiple views
- Data displayed to the user may differ from the actual data

Employee

The FAIR network consists of multiple facilities

An employee can have multiple roles

Any individual healthcare location, for example hospitals, clinics, or GGD's

One or more healthcare location(s) provide(s) access to one or more role(s)

Any employee role

Figure 32 Entity-Relationship Diagram Healthcare Process

The user processes of the FAIR-based framework, previously discussed in Sections 8.3.1 to 8.3.3, have been combined in a simplified representation of the complete system and is depicted in Figure 33. A more detailed representation of the various sub-processes can be found in Sections 8.3.1 to 8.3.3, which include a flowchart of each of these sub-processes in the context of the healthcare process and the aggregational statistics. More detailed information about the information flow in the context of the monitoring system is discussed in Section 8.5.2. More detailed information about the information flow in the context of data subjects having access to their own records is discussed in Section 8.5.6.

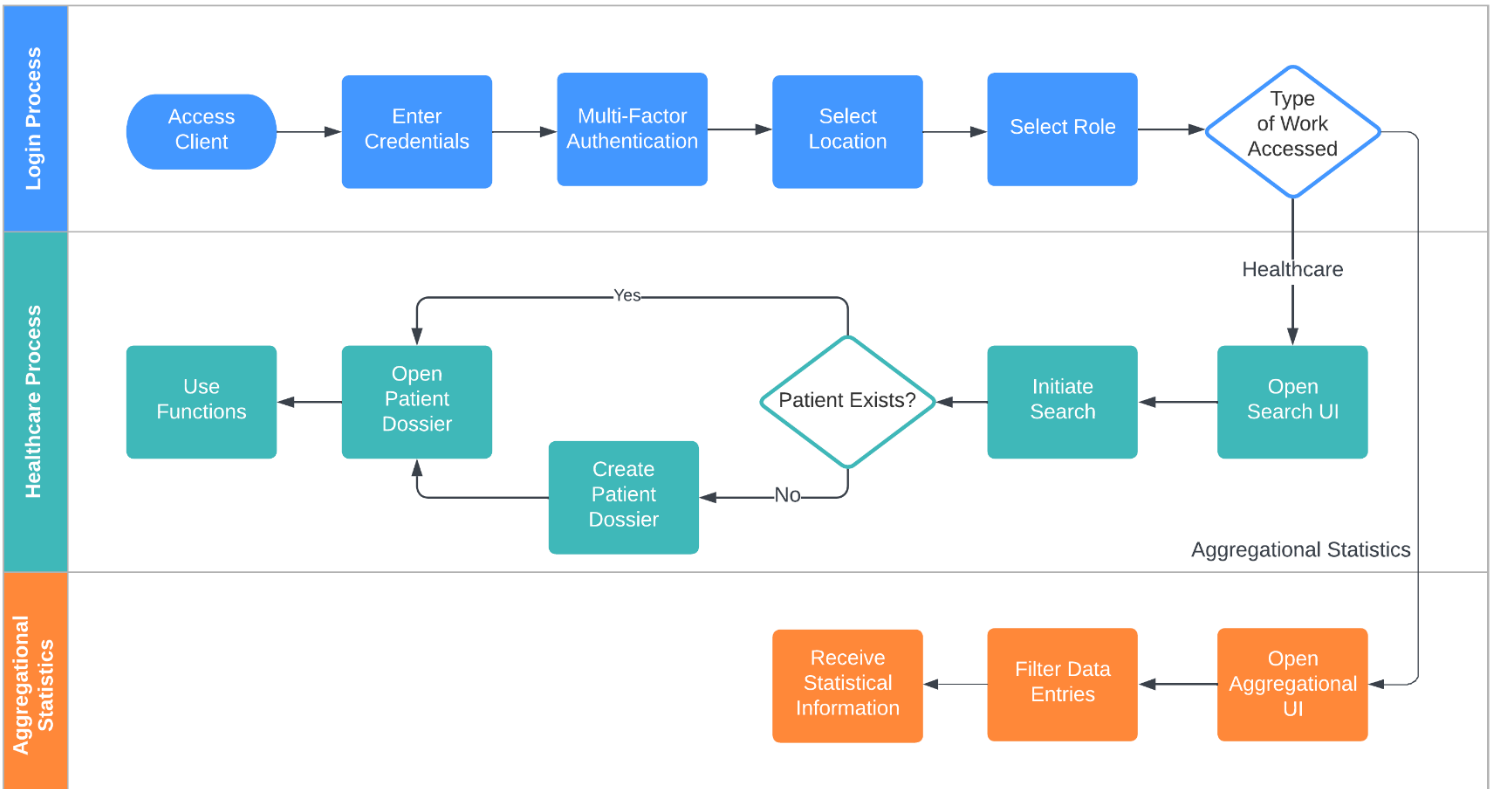


Figure 33 - Simplified representation of the FAIR-based Framework

8.4.1 FAIR Data Ownership and GDPR-compliant backups

This section will describe a backup solution that can be used in the healthcare sector, while still respecting the concepts of data visiting and data ownership at healthcare facilities. In addition to this, this solution is also designed to be GDPR compliant by being able to address the issue of changes based on data subject rights requiring technical or organizational measures to also be implemented in the backup system.

This solution will use the 3-2-1 backup strategy, referring to having three copies of data, stored on two different media, with at least one copy off-site. While other solutions can also be utilized, this should be seen as the minimum requirement. This form of redundancy ensures that data is unlikely to be lost due to either hardware failures in your primary system or any event that would be able to affect the entire facility. This has become of greater importance due to the significant rise of attacks on healthcare facilities as discussed in Section 7.3.1, where a ransomware attack would be able to encrypt all data which without a backup solution would be entirely lost.

The storage of data is duplicated over three elements, which are the primary system in which data is actively being processed, a write-only data storage solution at the same facility as the primary system, and a write-only data storage solution at an off-site location. For the off-site location, a specialized governmental facility such as one of the four centralized data centers intended for use by the central government could be considered. These facilities are already set up to securely store sensitive data and would not require significant investments. This has been visualized in Figure 34.

The transfer of data to these storage solutions can be done in multiple ways, although the minimal requirements are that the data is encrypted both in transit and at the facility. With an incremental back-up instead of a full backup, the amount of data that needs to be transferred should be limited enough to allow for the real-time transfer of data instead of scheduled transfers. This ensures that the data that is transferred is most up-to-date. Another advantage of this approach is that instead of having the database reflect a certain point in time, you could instead create a list of changes that could be applied to the database. This eliminates the need for a backup of any certain date and prevents the loss of data due to mistakes entirely.

It should be noted however that while the transfer of data to the on-site backup solution should not realistically be bandwidth limited, the same cannot be said for the off-site backup solution which requires an internet connection. Realistically, no system of this kind should be able to exceed the current levels of bandwidth but in locations where this does apply, the transfer could be done via a physical method where a data medium is transferred to the off-site solution based on a certain schedule. This carries the risk that, if something were to ever happen to the facility, the data between two cycles is likely lost.

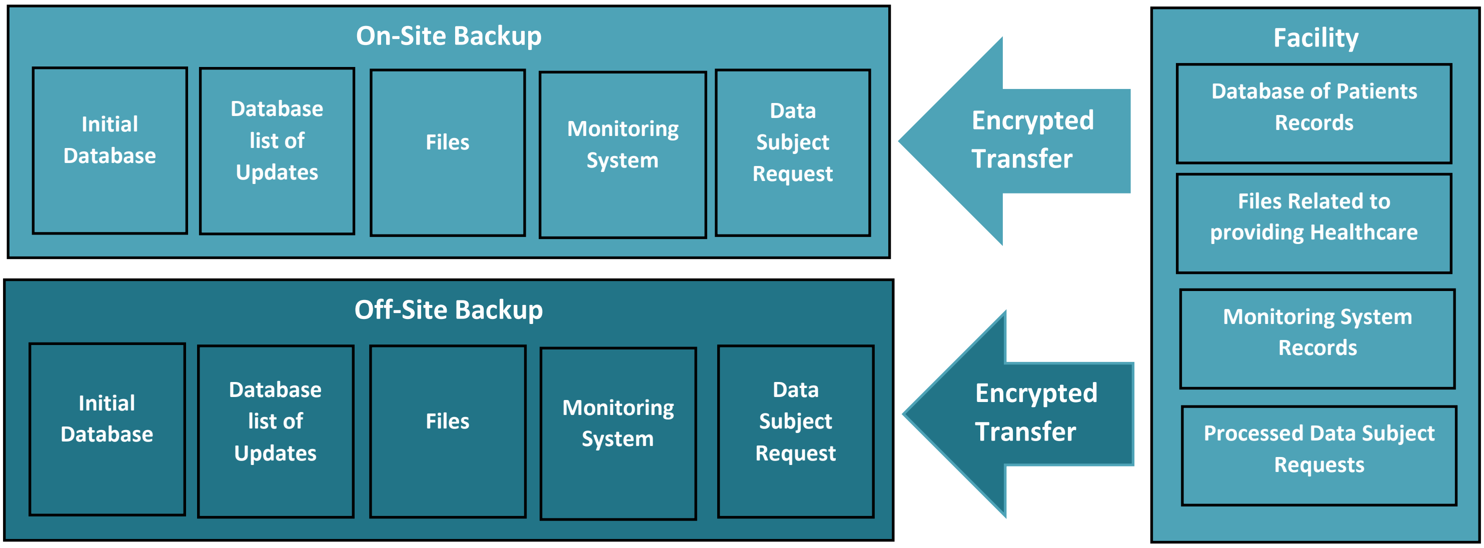
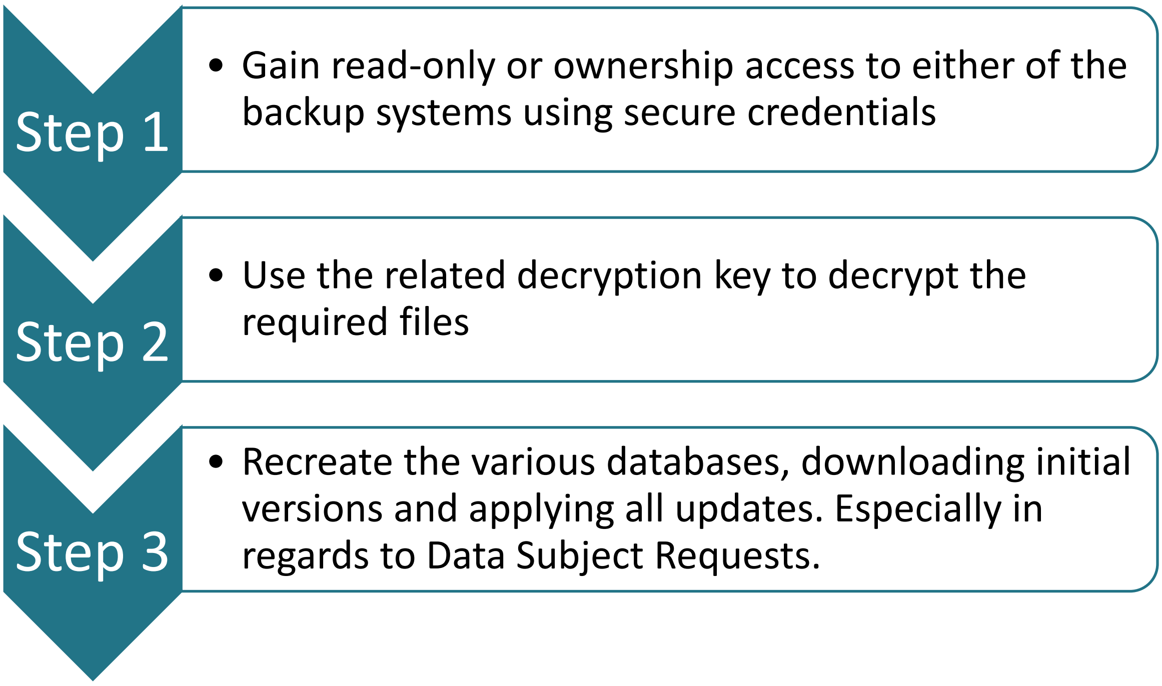


Figure 34 - Backup Approach

The process of restoring data from backups necessitates either read-only or ownership access to the dataset, which is initially write-only to prevent unauthorized access. Additionally, a decryption key is required to access the actual data in its unencrypted form as the data is encrypted to increase security even further. This process has been depicted in Figure 35.



Step 1

- Gain read-only or ownership access to either of the backup systems using secure credentials

Step 2

- Use the related decryption key to decrypt the required files

Step 3

- Recreate the various databases, downloading initial versions and applying all updates. Especially in regards to Data Subject Requests.

Figure 35 - Process to restore a back-up

To enhance security, additional credentials such as a hardware key can be used to restore backups. Two options are available, each with its own advantages and disadvantages. Storing the key on-site increases the risk of it being compromised in the event of an attack or a disaster. Conversely, storing it off-site removes the risk but raises concerns about the organization's data ownership. While digital keys can also be used, they increase the risk of the key being compromised even further. Other additions could be the use of a time-based lock with any request being able to be interrupted by an administrator-level user, although this would also increase the time required to restore a backup.

In the previous example, data has been stored in a write-only form which ensures that data can never be accessed without requiring additional action. This ensures that data will only be processed at the location itself, even while backups have been created. If a different approach is chosen, the key requirements are to ensure that data should never be able to be accessed by anyone outside of the organization the data belongs to. By following this requirement, the creation of a backup does not violate the principle of data visitation and ensures data ownership at the facility level.

8.5 System Design

This section presents the implementation of the various components within the proposed FAIR-based framework, which prioritizes data minimization and privacy by design and by default. Sections 8.5.1 to 8.5.4 detail the core components of any system, while Sections 8.5.5 and 8.5.6 discuss valuable additions that enhance the system's success or help it meet and surpass GDPR requirements, empowering data subjects with greater control over their data. It is important to note that this is merely the initial version of the proposed framework, which has been designed with continuous improvement and future expansion in mind. Additional components may be incorporated either as default components or as valuable additions to specific projects or sectors.

Section 8.5.1 describes the implementation of the access control system, which uses roles to regulate data and functionality access. Instead of employing a traditional permission matrix, users access assigned roles containing subsets of data and functionalities, considering the enhanced privacy, security, and usability this approach offers. In most cases, data access is limited to a single facility, with additional access provided only through the aggregational query system in Section 8.3.3, which does not expose any information.

Section 8.3.2 describes the implementation of the monitoring system, which collects various types of information from the login process, secure working environment, user search queries, and patient dossier activity. This data feeds into three distinct elements comprising the monitoring system: performance indicators, a Trained Behavior Engine, and a list of individuals of interest. The system's three functions are to store data for manual investigations, flag users for manual investigations, and automatically respond to instances that meet a certain threshold.

Section 8.5.3 describes the implementation of the search system, providing a detailed example of the user perspective for the system described in Section 8.3.1. The patient dossier in Section 8.3.2 is also further discussed.

Section 8.5.4 describes the Data Transformation System used throughout the entire framework. The main objective is to transition from an approach where users have access to a limited number of columns to one where even accessible columns have restricted information exposure. This section offers various examples, but numerous other implementations are possible. If implemented correctly, even unauthorized copying of returned information would not be able to result in a severe personal data breach.

Section 8.5.5 describes the implementation of a secure environment for users to ask and answer questions. In the GGD case study, it was found that employees often sought help from colleagues in insecure environments, compromising data security. The FAIR-based framework addresses this issue by providing a secure environment integrated into the platform, enabling organizations to maintain data control and implement measures to conceal sensitive information. Any attempt to bypass the integrated tool can be considered malicious and punished accordingly.

Section 8.5.6 discusses the implementation of a portal for data subjects. The GDPR grants data subjects specific rights regarding data ownership, as previously mentioned in Section 2.1.1. Organizations must respond to data subject requests within a certain timeframe or risk significant fines. This framework incorporates an online portal, accessible after identity verification, for data subjects to exercise their data ownership rights. This feature requires complete control over all processed data related to the data subject, enabling full automation. The portal offers data subjects a centralized location to exercise their rights, leading to time savings and reduced costs for organizations, as much of the system can be automated.

8.5.1 Access Control System

In the proposed FAIR-based framework, data should always be controlled and stored by governmental organizations or healthcare facilities to ensure data security. These entities bear full responsibility for managing this data, including securely providing access through data visiting and monitoring access by authorized users. Additionally, the entity must ensure the monitoring process is effective and take action when necessary. In the event that third-party organizations are required, those organizations should be provided with the required amount of data access but should not be controlling or storing any of this data.

The access control system of the proposed framework has been designed with crisis conditions in mind, such as the need to employ a large number of individuals lacking prior knowledge or connection to the healthcare sector. During the COVID-19 pandemic, the government extensively utilized external organizations to work within the same systems as healthcare facilities. In the event that external organizations are required, those organizations should be provided with the required amount of data access through data visitation, while data control, storage, and monitoring would still be managed by the data-managing entity.

Considering the high number of employees and the likelihood of adding new ones daily, data access could be granted to entire external organizations simultaneously, which would then provide low-level access to employees as needed. This also creates a hierarchy of data access, enabling the detection of organizational issues as they arise.

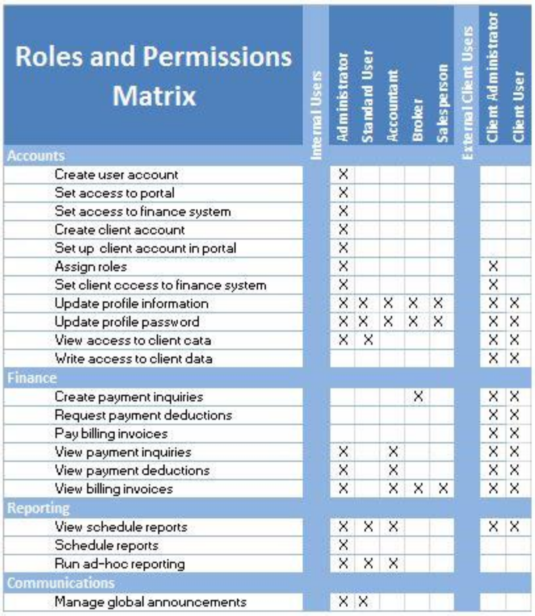
The following two sections will describe the conditions under which access will be granted to users and a description of the monitoring system used to monitor the processing process.

8.5.1.1 Roles

Since healthcare systems process highly confidential and personal data, access must be limited to authorized individuals who require data access for their duties. Each access instance should require valid user credentials. However, user credentials alone can be insecure and easily compromised through guessing, data breaches, or careless handling. To address this, a minimum two-factor authentication security level should be required, mandating users to approve access requests on a separate device.

Authorized platform users should not have unrestricted access to all database data. Instead, access should be limited to the data and functionality necessary for their specific roles. This practice is already commonly used in the healthcare sector and beyond, with access regulated through an authorization matrix or roles and permissions matrix (as shown in Figure 36).

Traditionally, higher-ranking and technical staff are granted greater access by default. An example of which can be seen in Figure 36, where an administrator has access to most functionalities. This, however, may violate the GDPR principle of privacy by design and by default as access to such an extent isn't necessarily required in the context of an employee's responsibilities. Thus, this architecture improves upon this system with user accounts and tiered access, where the user's access level is determined by their account-associated roles, as visualized in Figure 37.



Roles and Permissions Matrix	Internal Users	Administrator	Standard User	Accountant	Broker	Salesperson	External Client Users	Client Administrator	Client User
Accounts									
Create user account		X							
Set access to portal		X							
Set access to finance system		X							
Create client account		X							
Set up client account in portal		X							
Assign roles		X						X	
Set client ccccss to finance system		X						X	
Update profile information		X	X	X	X	X		X	X
Update profile password		X	X	X	X	X		X	X
View access to client cata		X	X					X	X
Write access to client data								X	X
Finance									
Create payment inquiries					X			X	X
Request payment deductions								X	X
Pay billing invoices								X	X
View payment inquiries		X		X				X	X
View payment deductions		X		X				X	X
View billing invoices		X		X	X	X		X	X
Reporting									
View schedule reports		X	X	X				X	X
Schedule reports		X							
Run ad-hoc reporting		X	X	X					
Communications									
Manage global announcements		X	X						

Figure 36 - Roles and permissions matrix

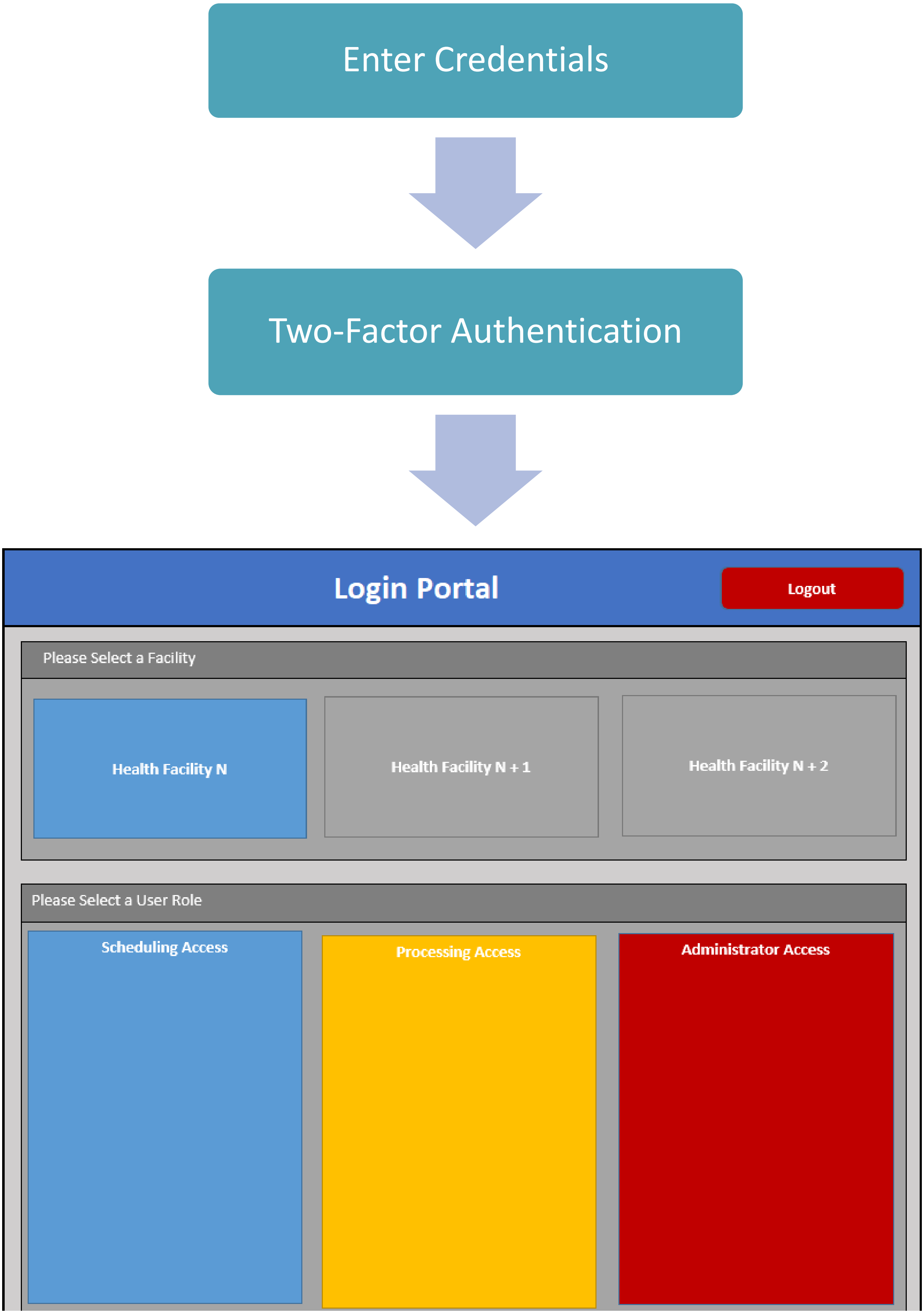


Figure 37 - User Login Procedure with User Roles

The first and second steps of this process resemble a traditional system, requiring users to enter credentials and verify their identity using a two-factor authentication method. The third step differs significantly, with the user being presented with a new page upon successful authentication. On this page, the user selects a specific health facility with granted access, limiting the current instance to data from that facility. The user is also presented with certain user roles based on their provided access level. Combining this approach with the logging system discussed in the next section allows for monitoring and tracking when an individual accesses a role with higher-level access or functionalities.

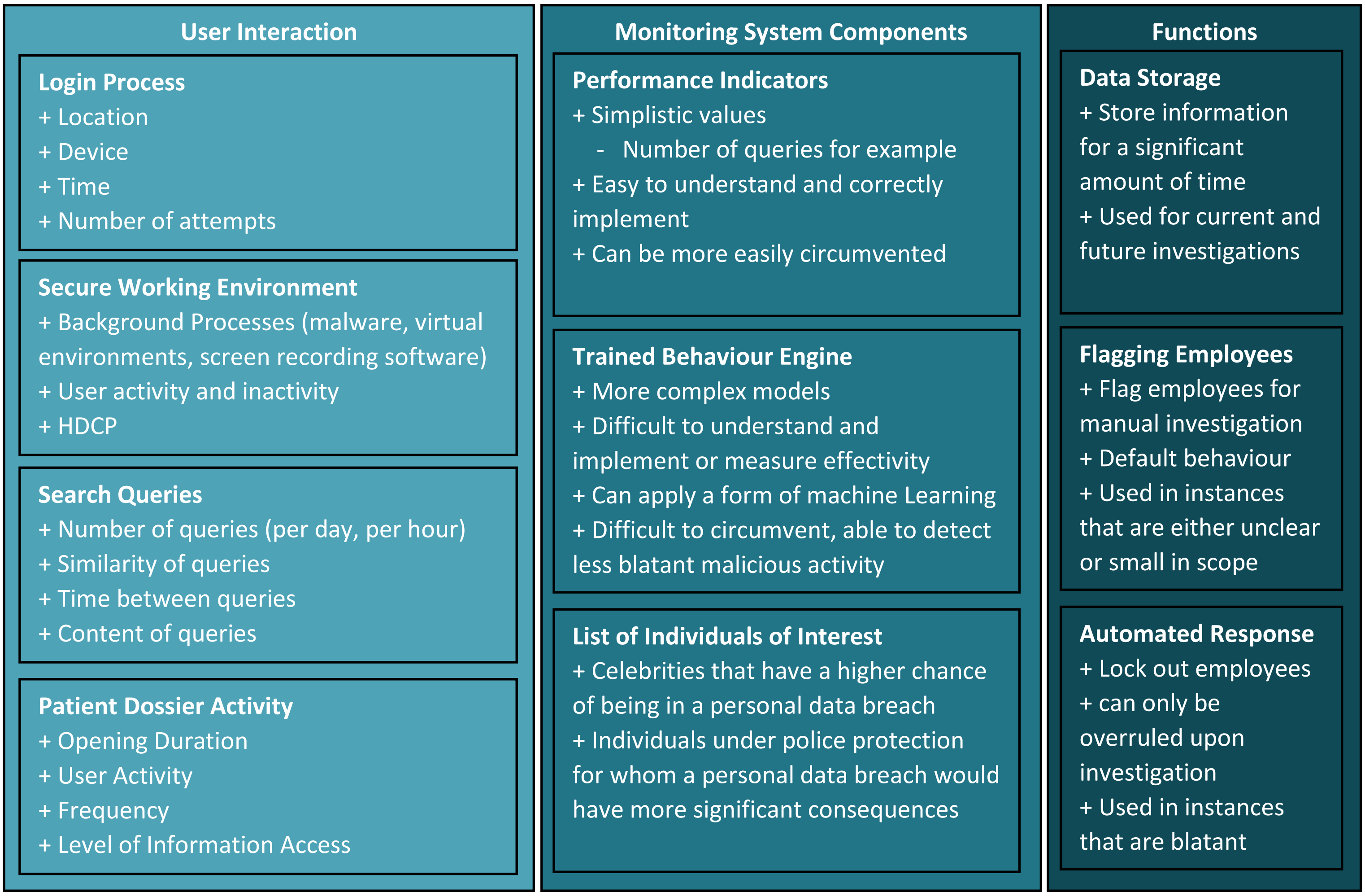
For example, a manager may have access to different roles with varying functionality levels depending on specific task requirements. Higher access levels could be separated into multiple roles with restricted functionality. One such functionality is record deletion, which could be included in a separate role from other functions. A technical developer would only have access to internal code and test cases necessary to verify system functionality, rather than unrestricted data access. While this is a logical restriction, it should still be stated as Mustafa indicated that this still occurs in governmental projects, including projects he was working on [191]. A call center worker would only have access to data and functionality necessary for their role, requiring login based on the roles their user account allows them to access.

Additionally, this proposed framework also addresses the issue of individuals receiving excessive access or retaining access after it is no longer needed. Which the case study found to be a prevalent issue at both the GGD and the various external organizations. To tackle this problem, access should be granted on a time-limited basis, avoiding extended periods of access. For instance, granting access for a year does not address the underlying reasons for implementing time-based access. A safer and more effective approach involves continuously granting access to employees as needed and revoking permissions when they are no longer necessary. This time-based system ensures that any mistakes will eventually be resolved without intervention. Furthermore, if a user has not logged in for a specified period, their access should be automatically revoked.

Even with the proposed data access regulation scheme in place, certain functionalities may pose inherent risks. In such cases, access to these functions should be limited based on an additional security system, such as a one-time approval system that only a select group of high-level employees can authorize. For example, exporting personal data from the system warrants this level of security. While the architecture aims to keep data within the system, situations may arise where data export is necessary. However, this introduces a significant security risk, and as a result, strict security measures must be enforced.

8.5.2 Monitoring

The monitoring system used in this architecture does not significantly differ from monitoring systems that could be implemented in other systems. However, as this architecture has focused on maximizing control over data by ensuring that all data is contained in the system, it becomes possible to record nearly all interactions with the healthcare data. Therefore, while the monitoring system itself isn’t new, this approach to data management makes the system more effective. An overview of the specified monitoring system Is depicted in Figure 38.



User Interaction

Login Process
+ Location
+ Device
+ Time
+ Number of attempts

Secure Working Environment
+ Background Processes (malware, virtual environments, screen recording software)
+ User activity and inactivity
+ HDCP

Search Queries
+ Number of queries (per day, per hour)
+ Similarity of queries
+ Time between queries
+ Content of queries

Patient Dossier Activity
+ Opening Duration
+ User Activity
+ Frequency
+ Level of Information Access

Monitoring System Components

Performance Indicators
+ Simplistic values
 - Number of queries for example
+ Easy to understand and correctly implement
+ Can be more easily circumvented

Trained Behaviour Engine
+ More complex models
+ Difficult to understand and implement or measure effectivity
+ Can apply a form of machine Learning
+ Difficult to circumvent, able to detect less blatant malicious activity

List of Individuals of Interest
+ Celebrities that have a higher chance of being in a personal data breach
+ Individuals under police protection for whom a personal data breach would have more significant consequences

Functions

Data Storage
+ Store information for a significant amount of time
+ Used for current and future investigations

Flagging Employees
+ Flag employees for manual investigation
+ Default behaviour
+ Used in instances that are either unclear or small in scope

Automated Response
+ Lock out employees
+ can only be overruled upon investigation
+ Used in instances that are blatant

Figure 38 - Overview of the Monitoring System

The monitoring system relies on data gathered from four elements where the user interacts with the platform. These are the login process, a secure working environment, user queries, and the user’s activity in patient dossiers. The secure working environment refers to a client-based approach instead of a browser-based platform, which allows for more data to be gathered beyond what is possible in a browser environment. The user interaction aspect of the monitoring system will be discussed in more detail in Sections 8.5.2.1 to 8.5.2.4.

The data gathered by the monitoring system through user activity is used by three different elements that all aim to detect malicious activity. The first element consists of performance indicators measuring simplistic values which are easy to understand and implement but may be easily circumvented. The second element is a trained behavior engine that applies more complex models which are difficult to understand and implement but are also difficult to circumvent and would be able to detect less blatant malicious activity. The third element is a list of individuals of interest, for example, celebrities and people under police protection, that will flag any employee that interacts with one of these individuals. These elements will be discussed in more detail in Section 8.5.2.5.

The monitoring system itself has three main functions based on the information gathered by the components of the monitoring system. The first function is to store information for a significant amount of time, measured in years instead of months, to support both manual investigation and any future automatic investigation with modern models being applied to historical data. The second function is to be able to flag employees that display possible malicious activity, which can then be manually investigated using the previously mentioned storage of data. The third function is to take automatic action, which in most cases would be to lock the user out of the system if the malicious activity exceeds a certain threshold. These functions will be discussed in more detail in Section 8.5.2.6.

8.5.2.1 Login Process

The login process is divided into the stage where a user enters his credentials and the stage where the user verifies their login attempt using a second method of authentication. For the credentials stage, the most valuable information is the location of the user, the device that is used, the time that the user logs in, and the number of login attempts. If the location of the user differs from the normal location, the user may be in a different location but it may also be a possible indication that the user's credentials have been compromised. If the device of the user differs from the normal device, the user may have a new device but this shouldn’t happen often or not with more devices than a laptop and a home desktop. If the time of the login attempt differs from the user's normal working hours, or the time differs from when the health facility would even be active, this is a significant indication that the user is either compromised or may be acting maliciously. Finally, the number of attempts may also be an indication that credentials are compromised or the user simply forgot his password. In both instances, the user could be requested to change their password for additional security.

For the second method of the authentication stage, the most valuable information is the device that is used and the location of the user. If the device of the user differs from the normal device, the user may have a new device but this shouldn’t happen often as it would only occur in the event of the user losing access to their 2FA device which is likely going to be their phone. If the location of the user differs from the normal location, the user may be in a different location or the user may be compromised although this is significantly more unlikely given that this would require either the loss of a physical device or some form of social engineering to be able to use a different device as the second method of authentication. The location can also be used by comparing the location of the second method of authentication and the location of the user, which should match. If they do not, it is unlikely that the user is the one instigating the login attempt. This is especially important as a user that is not careful could have their credentials compromised and then authorize an authentication request from someone else entirely by mistake.

8.5.2.2 Secure working environment

The secure working environment aims to detect suspicious behavior and prevent personal data from being able to be extracted in the most common ways. It is technologically impossible to detect all avenues of attack or completely prevent personal data from being extracted, however. This section discusses the properties of a client created for the health portal, instead of being limited to the properties of a web-based client, including which attacks are mitigated through these properties. The final paragraphs discuss areas that are unable to be completely prevented by this element, requiring other elements of the FAIR-based framework to be used.

As the secure working environment is an installed client on the user's computer, the client would be able to monitor background programs and detect the environment in which it is run. The detection of background programs could detect the use of keyloggers, screen capture tools, and potential malware which could compromise the security of the data. The detection of a strange environment such as a virtual machine is not an indication of the malicious activity itself, but it could be flagged for further investigation as this situation will represent a fraction of the total user count and could potentially be used to circumvent security measures.

The client also allows for the recording of additional information related to user activity which could be used to both detect malicious activity as well as tracking performance and functioning as a way to log users out of their sessions based on a period of inactivity. There is a significant amount of possible customization using this approach but one example is to flag users that copy information from the client by detecting data that is being highlighted or by detecting the copy command.

In addition to monitoring, the client could also prevent more simplistic attacks from working by implementing existing standards aimed to prevent the copying of digital content. Using 'High-bandwidth Digital Content Protection', many screen recording software, as well as capture devices, are unable to make a copy of any digital content. The screen would be replaced by a black image, preventing any information from becoming compromised.

However, these methods are only able to improve the detection of malicious activity and prevent some attacks from working. As such, it is unable to guarantee security. Yet, this does not mean that these methods are unable to significantly increase security. While these methods can be circumvented through something as simple as taking a picture from a mobile phone or by using any screen recording software or capture device that can circumvent the HDCP standard, it can detect and deter more opportunistic criminals.

8.5.2.3 Search Queries

User search queries can be used to monitor user interaction in multiple ways. While the information offered by a query is more limited than the information offered in a patient dossier, information is still being exposed to the user. Which in most instances can still have significant value, especially in the use of targeted phishing attacks based on the data subject's health. The main sources of information in this element are the number of queries, the similarity of queries, the time between queries, and the content of queries. This information can then be compared with what would be expected activity from a normal user, to determine if there is any indication of malicious activity.

The number of queries conducted by a user can be used to detect the most blatant attempts of extracting information. The number of queries could be measured per day, per hour, or any other metric and then compared to the value of a normal employee. If this value is significantly exceeded, this could indicate the user is not using queries in support of his responsibility and instead uses them to extract information. The GGD case offers a good example of this, as discussed in Section 5.5.3, a user accessed hundreds of patient records which likely require hundreds of queries in a very short period. Measuring this metric would have immediately detected this blatant malicious activity.

The similarity of queries conducted by a user can be used as a mitigating factor for the previous metric, to distinguish more efficient users from less efficient users using the system for malicious activities. Given the increased query requirements in this FAIR-based framework, which will be discussed in Section 8.5.3, it is likely that the number of queries required by a user will increase. A normal workflow would show the user continuously expanding on an initial query, which the monitoring system could record. If there is a high number of queries, with a low level of similarity between them, this would indicate that the user is not acting in support of this responsibility and should be flagged for malicious activity.

The time between queries conducted by a user is another metric that can be used to support the previous ones. In a normal workflow, there should be some time between queries as the user interacts with a new data subject. In the workflow of this FAIR-based framework, this metric should show relatively little time between queries that deepen initial queries, while there should be relatively more time between completely different queries. If this is not the case, this would indicate that the user is not acting in support of this responsibility and should be flagged for malicious activity.

The content of queries conducted by a user is another metric, which is especially valuable for the list of individuals of interest. Instead of measuring the similarity between queries, this metric looks at the actual queries, including which information is being returned to the user. This may differ from normal behavior if there is a significant increase in the number of individuals of interest, or if a specific group of people is targeted, such as older people in wealthier regions of the Netherlands which would be the prime victims of malicious activity. It should be noted, however, that in a healthcare system, a significant number of people would belong to this group as older people are generally at a heightened risk compared to younger ones. The activity needs to differ from regular behavior to be considered a possible malicious activity.

While it is likely that significantly more metrics can be created related to user activity in search queries, these are beyond the scope of this thesis. The purpose of this section is to present a practical implementation of this concept, which can always be extended with additional metrics.

8.5.2.4 Patient Dossier Activity

The information in patient dossiers is of additional value compared to the more limited value offered by a query and therefore requires additional monitoring. Given the increased value of information, malicious activity detected via this component should also be considered as having a higher value. The main sources of information in this element are the opening duration of the patient dossier, user activity within it, the frequency of patient dossiers being opened, and the amount of exposed information. This information can then be compared with what would be expected activity from a normal user, to determine if there is any indication of malicious activity.

The opening duration of the patient dossier is the first metric that will be discussed in this section. In the context of malicious activity, this duration would likely be low as the user requires little time to make a copy of the information in the patient dossier. If the opening duration does not match the normal duration of other employees, this could be an indication of malicious behavior. Depending on the way the transformation feature discussed in Section 8.5.4 is implemented, this could increase the duration patient dossiers are opened as more user interaction is required to access information.

Another metric is the user activity in a patient dossier. This activity would likely differ from the normal workflow when it is accessed in the context of malicious activity. Although this metric has significant overlap with the monitoring of user activity in the secure working environment, which also monitors which fields are interacted with, or when the copy command is used for example.

The frequency in which a user opens patient dossiers is another metric that can be used, although it shares some overlap with the metric for the opening duration of patient dossiers. The previous example where a user accessed hundreds of patient records in a short period would result in a very high frequency and likely a low opening duration. However, a lower frequency that is still higher than normal behavior could display a normal opening duration with a high frequency that would otherwise not be detected as possible malicious activity.

The amount of information that is exposed is another metric that can be used, which offers significant value if one of the later methods of data transformation as discussed in Section 8.5.4 is used. As data is locked behind functions that require user activity to expose them, the monitoring system is offered another avenue for collecting data which also makes it possible to know exactly which data was accessed. If this differs from regular behavior, this could be marked as possible malicious activity.

While it is likely that significantly more metrics can be created related to user activity in search queries, these are beyond the scope of this thesis. The purpose of this section is to present a practical implementation of this concept, which can always be extended with additional metrics.

8.5.2.5 Components

The first component of the monitoring system is a series of performance indicators, based on the metrics discussed in the previous sections. These are simplistic values that are easy to understand and can easily be implemented in any such system. While they may have significant value, they are also easy to circumvent and only able to detect the most blatant malicious activity. Making this component has some value but requires additional components to be fully effective.

This problem is addressed through the second component, the trained behavior engine. The data input for this component remains the same, which is then processed into more complex models. The downsides of which are that the models become difficult to understand and implement, nor is it easy to measure the effectiveness of the approach. This could even be achieved through the use of black models that have learned to classify suspicious behavior but are unable to be easily explained by system developers involved with the system. This approach is difficult to circumvent as even with information from the system itself, it’s still unclear what is being detected, making it possible to detect even less blatant malicious activity.

The last component of the monitoring system is a list of individuals of interest which is already currently used in the healthcare sector. The purpose of which is to prevent users from looking up individuals of interest if they do not require this in the context of their responsibilities. While this does not entirely prevent his behavior, as these events still occur in the healthcare sector as discussed in Section 7.5.3.2, the significantly increased probability of being caught should reduce it significantly. This concept would work best with a centralized list created by the Dutch government, which could then be re-used across all layers of government as this problem is not exclusive to the healthcare sector. If such a list does not exist yet, it could be expanded from partial lists currently in existence until the creation of such a centralized list.

While the focus of the monitoring system is to prevent personal data breaches from occurring, the individuals on this list could be significantly more affected by a personal data breach than others. While the release of contact information such as a phone number could be seen as an annoyance for many people, the release of address details of celebrities to a stalker of the address details of individuals under police protection to criminals could be a serious threat to their life. Given the simplicity of this approach and the potential value, this is a required inclusion in any such system.

8.5.2.6 Functions

It is of critical importance that the data recorded via the monitoring system is stored for a significant period, measured in years instead of months. There are two main purposes to this, the first of which is to make it possible to manually or automatically investigate users and the second of which is to act as a deterrent the first one is to act as a deterrent as all activity is stored for a significant amount of time so that even if a user's malicious activity is not detected initially, this is no guarantee that it will never be detected. In the future, better models could be applied to the data set to detect malicious activity that previously went undetected.

If monitoring data is not stored for an appropriate time, it is impossible to accurately determine the scope of any personal data breach, which would result in many data subjects becoming affected without being made aware of the fact. Something which likely occurred during the GGD case, as discussed in Section 5.5.3, where monitoring information was stored for only a single month, making any investigation into the true scale of the personal data breach impossible.

Given the nature of the data that is stored in a monitoring system, and its critical value in ensuring the security of the data processing process, it is unlikely that any restrictions under the GDPR would prohibit the storage of this data for a longer period. Especially due to its continued value long after the user activity as better models could be applied to historical data in the future, which implies that the data continues to remain of use and therefore should be deleted.

The primary function of the monitoring system is to flag users if there is a suspicion of malicious activity. This term refers to adding a mark to a user, indicating that the user should be subject to manual review. The manual review process would consist of a human carefully investigating the user, either in the specific instance for which the user was flagged or in a more general situation, to determine if the user was indeed acting maliciously and what the scale of the personal data breach is. In the event of a data breach, the event would be reported to the Autoriteit Persoonsgegevens.

Using this approach, the probability of an employee being unfairly sanctioned is very low. However, given the time required for manual review, it requires a significant number of employees to perform these investigations. To limit this, an addition could be made to the flagging system to indicate various priorities which address the users with the highest probability of malicious behavior first. However, this does not address the fact that if a user was acting maliciously, continued access would only increase the probability of another personal data breach or increase the scope of the current one.

To address this, the final function of the monitoring system is to employ automatic responses. Which would involve locking the user out of any access to the system, which can only be restored upon manual review. While the exact conditions of such a system are highly customizable, in general, such a system would be employed in instances where there is a significant indication of malicious activity, or the most blatant attempt to commit malicious activity. In any such system, there should be a balance between restricting access, which could affect normal employees and reduce efficiency on one side, and reducing the probability and scope of a personal data breach with highly sensitive information.

8.5.3 Search functionality

As discussed in Section 2.2.1, digital resources should be easily found and identified by both humans and computers. To achieve this, digital resources must be indexed by search engines and other discovery tools using a unique identifier, described with detailed metadata, and properly linked and labeled. Which extends to the design of both the search process and the application or website itself. However, in the context of supporting a healthcare process, a search engine that has implemented this principle may provide access to too much information, beyond what Is required for a certain role.

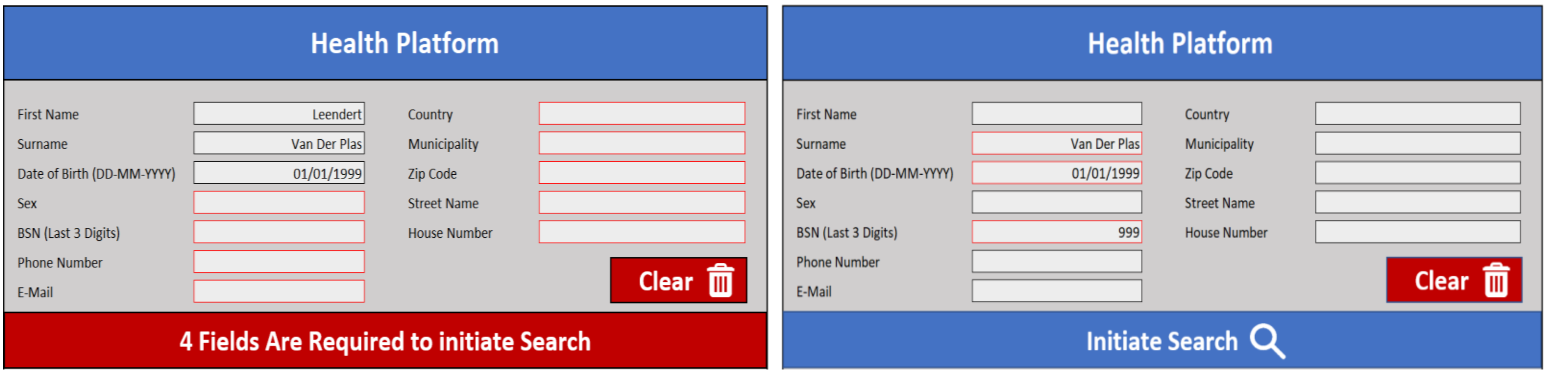
To address this issue, the proposed architecture will include improvements to the search functionality that support data minimization and privacy by default and by design. These improvements will increase the amount of information required to initiate a search and limit the information that is initially returned. This is further improved through a balance of the amount of information that is provided and the amount of information that first requires interaction.

8.5.3.1 Increased requirements for initiating a search query

If the data in a database is accessible via a very general search query, this poses a potential security risk as healthcare data is highly confidential and valuable. Even data displayed in an initial search could be sensitive and allowing such broad access would violate the principles of privacy by default and by design. Additionally, this approach allows for a simple attack that is difficult to detect, as recording the search results would not differ from normal behavior and would not be flagged by automated or manual controls. As accessing this data is both required and justified, this should be addressed in a manner that does not prevent employees from carrying out their job.

To address this issue, the search functionality in the FAIR-based framework will require a certain amount of information to be entered before initiating a search query to limit the number of results that are returned to the user. There are two possible approaches, however, both aim to prevent search queries from being too generic and exposing information unnecessarily.

The first approach is to set either a minimum number of fields that need to be entered or certain specific fields that need to be entered or a combination of both before initiating a search. For example, the search engine could require the user to enter four fields in order to be able to initiate a search (Figure 39 - Left). Or when specific fields are required, the search engine could require the user to enter both a surname, a date of birth, and the last three digits of a BSN (Figure 39 - Right). Both approaches could also be combined to require both a minimal number of fields and certain specific fields to be filled before initiating a search query.



Health Platform

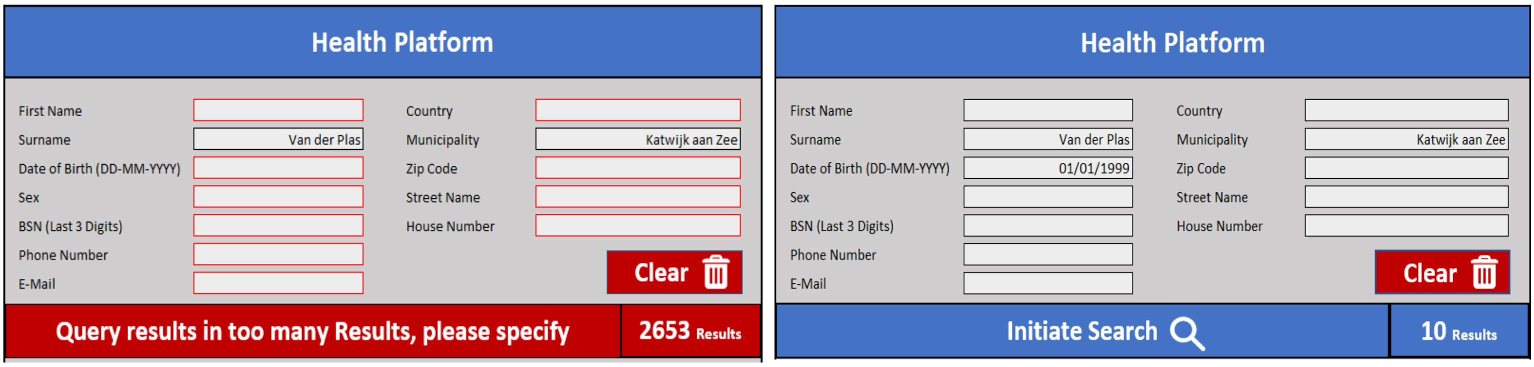
First Name	Leendert	Country	
Surname	Van Der Plas	Municipality	
Date of Birth (DD-MM-YYYY)	01/01/1999	Zip Code	
Sex		Street Name	
BSN (Last 3 Digits)		House Number	
Phone Number			
E-Mail			

Health Platform

First Name		Country	
Surname	Van Der Plas	Municipality	
Date of Birth (DD-MM-YYYY)	01/01/1999	Zip Code	
Sex		Street Name	
BSN (Last 3 Digits)	999	House Number	
Phone Number			
E-Mail			

Figure 39 - Search Query based on Fields
Number of fields on the left VS. Specific fields on the Right

The second approach is similar to the previous approach but instead of determining the number of fields or requiring certain fields to be filled in, this approach determines a search query to be too generic by evaluating the number of results that would be returned. Only if the number of results that would be returned is below a certain threshold would the results be displayed to the user. For example, the search query for data subjects with the surname of Van der Plas in Katwijk aan Zee would be a query that is too generic given the number of results that would return (Figure 40 - Left). The addition of a date of birth significantly reduces the number of results to an acceptable level where results may be returned to the user (Figure 40 - Right).



Health Platform

First Name		Country	
Surname	Van der Plas	Municipality	Katwijk aan Zee
Date of Birth (DD-MM-YYYY)		Zip Code	
Sex		Street Name	
BSN (Last 3 Digits)		House Number	
Phone Number			
E-Mail			

Query results in too many Results, please specify

2653 Results

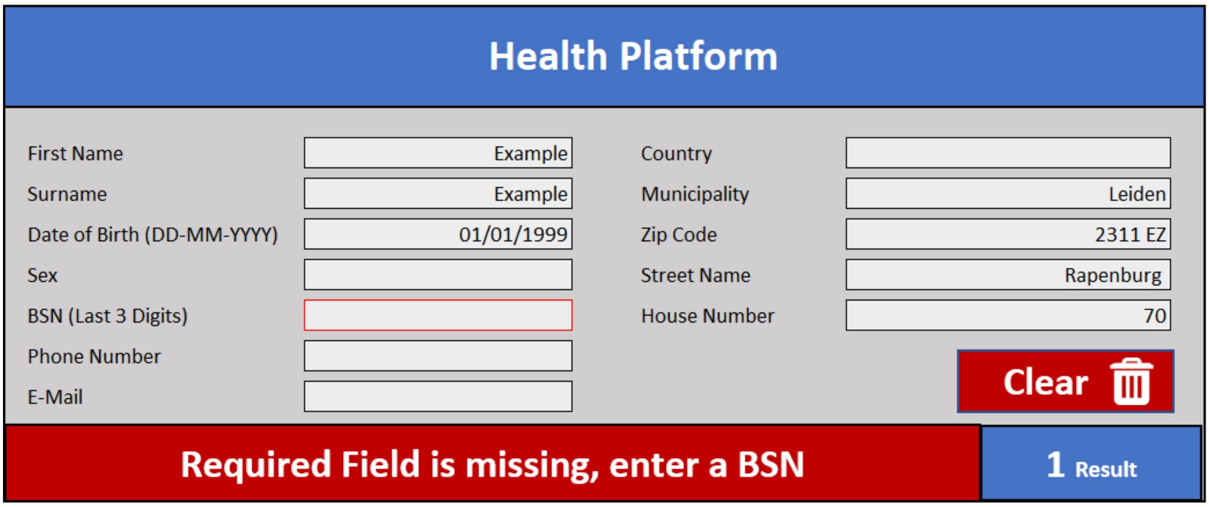
Health Platform

First Name		Country	
Surname	Van der Plas	Municipality	Katwijk aan Zee
Date of Birth (DD-MM-YYYY)	01/01/1999	Zip Code	
Sex		Street Name	
BSN (Last 3 Digits)		House Number	
Phone Number			
E-Mail			

10 Results

Figure 40 - Search Query Based on the Number of Results
Generic Query on the left VS. Specific Query on the Right

This approach places fewer restrictions on the user, but may also be combined with the previous approach to require specific fields, with additional fields being required depending on the number of results the query would return. This may be a requirement as the amount of information available about data subjects may still allow an employee to access records, they have no reason to access. For example, all information except for their BSN, their phone number, and e-mail address is likely to be available for any given person of interest in the Netherlands. Given this information, it is possible to construct specific queries that this approach would allow. Therefore, it may be advisable to require the entering of certain fields such as a partial BSN to prevent this (Figure 41Figure 40).



Health Platform

First Name	Example	Country	
Surname	Example	Municipality	Leiden
Date of Birth (DD-MM-YYYY)	01/01/1999	Zip Code	2311 EZ
Sex		Street Name	Rapenburg
BSN (Last 3 Digits)		House Number	70
Phone Number			
E-Mail			

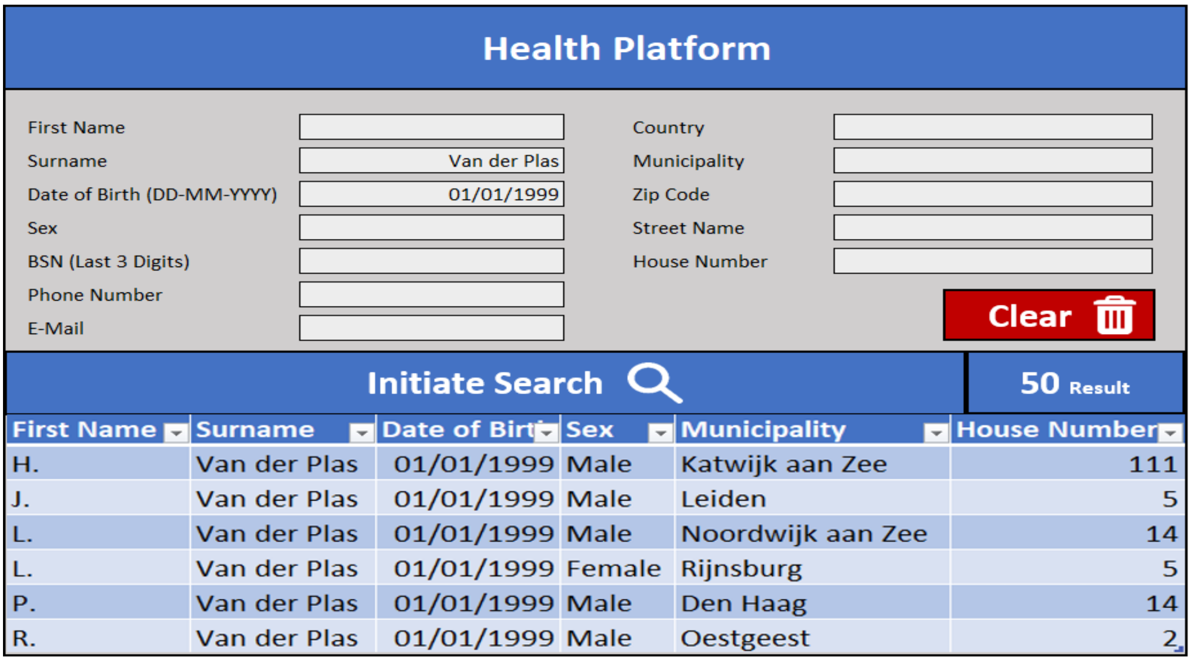
Required Field is missing, enter a BSN

1 Result

Figure 41 - Search Query Based on the Number of Results & Specific Fields

8.5.3.2 Limiting the amount of information returned by queries

Instead of using the number of results as the condition for which records may be returned, it can also be used to vary the amount of information returned to the user. Using this approach, different levels of information would be returned to the user, depending on the number of search results. Generic queries with many results would return only the most basic information for each entry (Figure 42), while a more specific query with a limited number of results would show more detailed information (Figure 43).

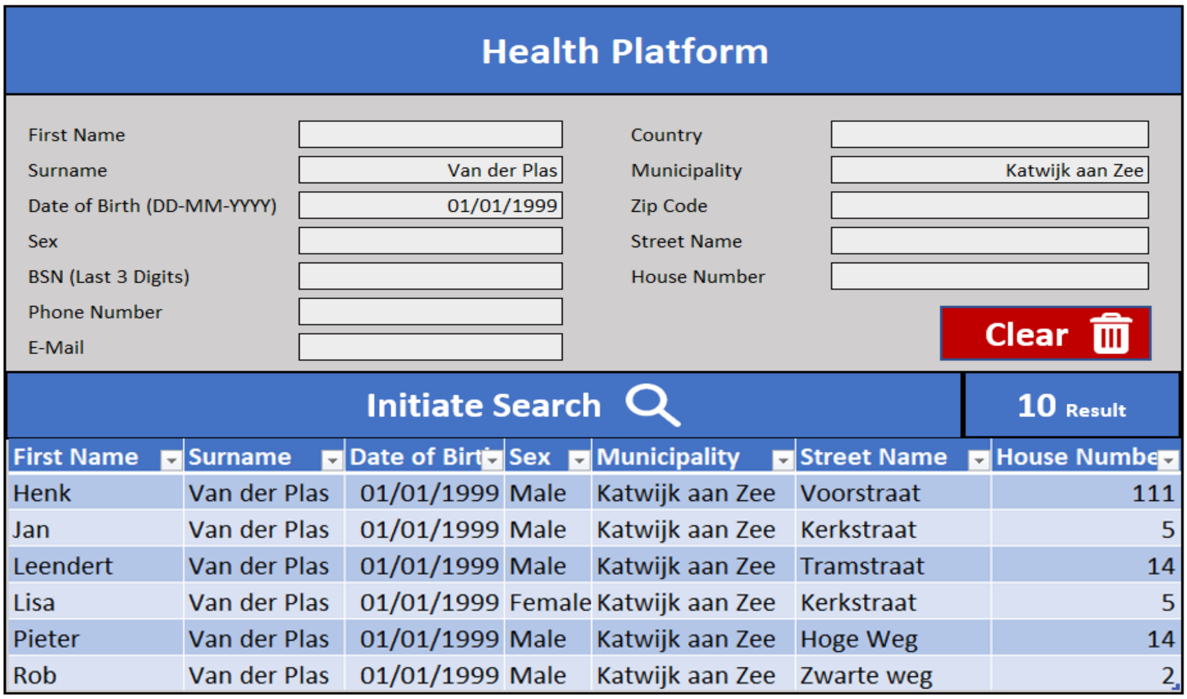


Health Platform

First Name		Country	
Surname	Van der Plas	Municipality	
Date of Birth (DD-MM-YYYY)	01/01/1999	Zip Code	
Sex		Street Name	
BSN (Last 3 Digits)		House Number	
Phone Number			
E-Mail			

First Name	Surname	Date of Birt	Sex	Municipality	House Number
H.	Van der Plas	01/01/1999	Male	Katwijk aan Zee	111
J.	Van der Plas	01/01/1999	Male	Leiden	5
L.	Van der Plas	01/01/1999	Male	Noordwijk aan Zee	14
L.	Van der Plas	01/01/1999	Female	Rijnsburg	5
P.	Van der Plas	01/01/1999	Male	Den Haag	14
R.	Van der Plas	01/01/1999	Male	Oestgeest	2

Figure 42 - Generic Search Query Results



Health Platform

First Name		Country	
Surname	Van der Plas	Municipality	Katwijk aan Zee
Date of Birth (DD-MM-YYYY)	01/01/1999	Zip Code	
Sex		Street Name	
BSN (Last 3 Digits)		House Number	
Phone Number			
E-Mail			

First Name	Surname	Date of Birt	Sex	Municipality	Street Name	House Numbe
Henk	Van der Plas	01/01/1999	Male	Katwijk aan Zee	Voorstraat	111
Jan	Van der Plas	01/01/1999	Male	Katwijk aan Zee	Kerkstraat	5
Leendert	Van der Plas	01/01/1999	Male	Katwijk aan Zee	Tramstraat	14
Lisa	Van der Plas	01/01/1999	Female	Katwijk aan Zee	Kerkstraat	5
Pieter	Van der Plas	01/01/1999	Male	Katwijk aan Zee	Hoge Weg	14
Rob	Van der Plas	01/01/1999	Male	Katwijk aan Zee	Zwarte weg	2

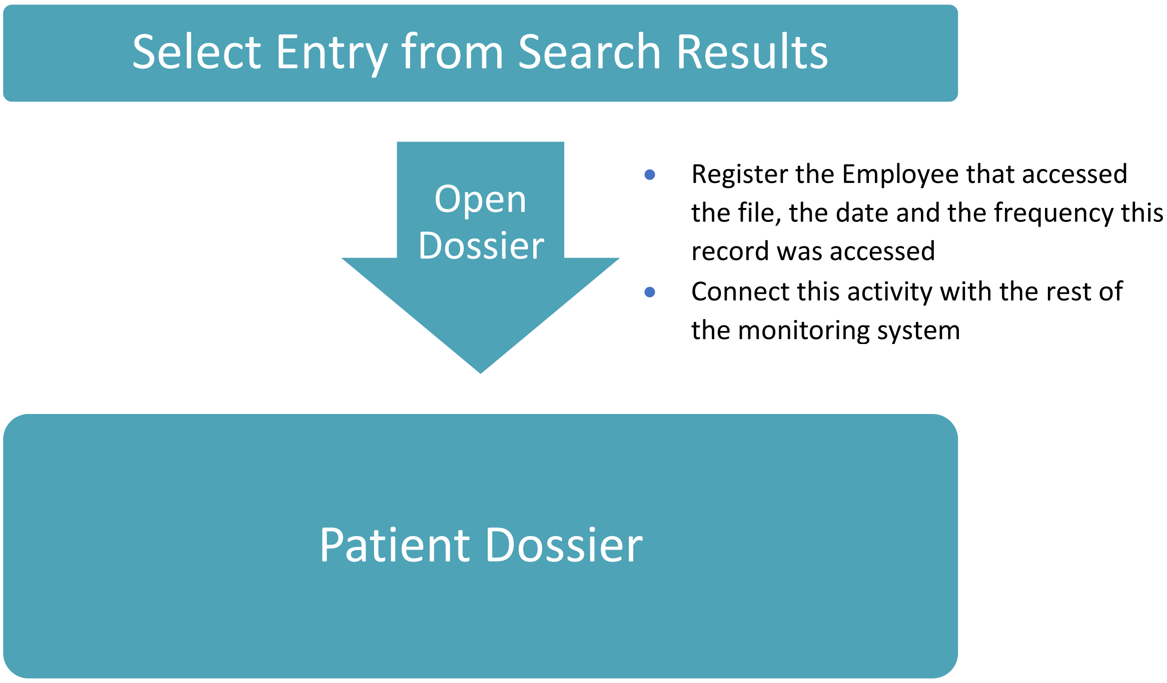
Figure 43 - Detailed Search Query Results

This approach provided an additional layer of protection for sensitive data, while not necessarily impacting the way search queries are used. As the goal of any such query is to find and access information from a specific data subject, nothing about this process is changed. The amount of information that would be returned to the user also matches this purpose, as more generic queries with a higher number of results would be easily filtered using non-sensitive information while more specific queries with a limited number of results would need specific and more detailed data field to be able to distinguish them and identify the correct data subject.

8.5.3.3 Patient Dossiers

In addition to addressing aspects of the search query function, adjustments can also be made to the balance of information that is returned to the user without further user interaction. By making use of a more detailed patient dossier upon accessing a sub-menu via an individual returned result, access can be provided to more sensitive information without compromising security. This approach is not necessarily new or even unique to the proposed FAIR-based-framework but is an important inclusion in the context of the secure processing of data as while patient dossiers have become the standard in the healthcare sector, systems can be designed without them or standards can be forgotten to be included when developing new systems.

This approach is inherently safer as access now requires user interaction, as depicted in Figure 44, which leads to there being a clear record of individuals that accessed specific information and reduces the effectiveness of the previous attack where search results would be recorded in some manner. This reduces the probability and scope of a personal data breach as it has become more difficult to access information without being monitored which would make individuals less likely to do so maliciously and the reduction of sensitive information being initially visible would reduce the scope of a personal breach.



- Register the Employee that accessed the file, the date and the frequency this record was accessed
- Connect this activity with the rest of the monitoring system

Figure 44 - Patient Dossier Access Process

8.5.4 Data Transformation

The processing of data often involves access to the full set of columns and values as specified in the processing description, however, this is not necessarily what a user requires. The actual requirement for the user is to be able to accomplish their task using the information from the data. Therefore, having access to the complete values stored in the database is not by definition a user requirement. By limiting the amount of data returned to the user to only what is strictly required, for example through transformations, menus, and functions, it is possible to make significant improvements to the principle of data minimization. This reduces the potential severity of personal data breaches and may also decrease the likelihood that confidential information will be leaked, as malicious actors would be less likely to accept the risk in return for the lower value.

It is worth noting that these methods can be combined and applied to any database, depending on the specific needs of the users and the functions of the system. By implementing these approaches, it is possible to improve data minimization and reduce the risks associated with data processing. Due to this, this method isn’t exclusive to a FAIR-based framework, although certain conditions have to be met to ensure that this approach is effective in preventing personal data breaches.

If the transformation is done on the computer of the user, while the original data is transmitted in full, then malicious actors would still be able to gain access to the original information. This would make this approach function only as a deterrent for less motivated malicious actors, while more experienced malicious actors would be able to extract the information directly from the transfer.

This concept of Data Transformation is illustrated in Table 20 to Table 25, which show different approaches to limiting the amount of data returned to users. Table 20 displays the full data entry for an individual, including all columns stored in the database, which is meant to be the benchmark to which these transformations will be compared. The columns themselves are based on the information that was included in the CoronIT system and most if not all of these columns are standard for any system used in the healthcare sector.

Table 21 demonstrates the removal of certain columns based on user needs, reducing the amount of high-value personal data and therefore the potential impact of a data breach. Although this technique is not limited to high-value personal data. It can and should be applied to all data that the user does not require.

Table 22 shows the transformation or limitation of data to display less information while still allowing for various checks. This is mainly related to checks of identity. As it is rarely required to use the entire value as a confirmation check, the user does not have any requirement to access the entire value. However, if other systems also use the final digits of a data field to perform verification checks, even this limited amount of information has significant value.

Table 23 introduces a simple change to increase security by making data fields visible only upon user action. This reduces the likelihood of a "screenshot attack" by requiring users to actively reveal sensitive information, such an example has been discussed in Section 5.5.3. Combined with a proper monitoring system, it would be significantly easier to distinguish malicious users from other users. This does come at a performance loss, with the exact loss needing to be determined in a system design study.

Table 24 illustrates a data processing approach where the user now asks the individual for information, which they then enter into the system and the system will respond with either a validation or denial. While the same amount of information is being shared, there is next to no risk of individuals processing data beyond their duties. The only time when information is processed is when the data subject can give consent. While users are still able to write down this information, it has become impossible to make a picture of screens to share data that way.

Table 25, like Table 24, presents a more radical approach to data processing, using functions to perform tasks without displaying personal data. For example, when sending an email, the user does not necessarily need to know the recipient's email address. By creating a function that enables the user to send the email to the correct recipient without displaying the data entry, it is possible to prevent the leak of personal information. This approach has the potential to significantly reduce the risks associated with data processing while still allowing users to perform necessary tasks.

The effectiveness of these concepts is enhanced when they are integrated into other parts of the architectural framework. For example, by connecting this system to the user's role (as discussed in Section 8.5.1.1), it is possible to provide different users with different levels of access. By connecting this system to the user activity tracking system, it becomes possible to report exactly which data has been accessed for each data subject, making it easier to distinguish between accidental mistakes and malicious activity (as discussed in Sections 8.5.2.3 and 8.5.2.4). Additionally, this system of data transformation can also be applied in the academic sharing aspect of the framework (discussed in Section 11.3), allowing the user to access different levels of information while still processing the actual data in the database.

Identifier	First Name	Surname	Date of Birth	Zip Code	House number	Street	Place of residence	Country	Linked GGD	Phone number	E-mail	BSN
23098283	Leendert	Van der Plas	01/01/1999	2311 EZ	70	Rapenburg	Leiden	Netherlands	Hollands Midden	0612345678	Leendert@leiden.nl	123456782

Table 20 - Database Entries 1

Identifier	First Name	Surname	Date of Birth	Zip Code	House number	Street	Place of residence	Country	Linked GGD	Phone number	E-mail	BSN
23098283	Leendert	Van der Plas	01/01/1999	2311 EZ	X	Rapenburg	Leiden	Netherlands	Hollands Midden	X	X	X

Table 21 - Database Entries 2 - Removing Columns

Identifier	First Name	Surname	Date of Birth	Zip Code	House number	Street	Place of residence	Country	Linked GGD	Phone number	E-mail	BSN
23098283	L.	Van der Plas	20-25	EZ	70	Rapenburg	Leiden	Netherlands	Hollands Midden	xxxxxxxx78	Leen....@leiden.nl	xxxxxx782

Table 22 - Database Entries 3 - Reducing Entries

Identifier	First Name	Surname	Date of Birth	Zip Code	House number	Street	Place of residence	Country	Linked GGD	Phone number	E-mail	BSN
23098283	Leendert	Van der Plas	01/01/1999	2311 EZ	70	Rapenburg	Leiden	Netherlands	Hollands Midden	Show	Show	Show

Table 23 - Database Entries 4 - Visibility Functions

Identifier	First Name	Surname	Date of Birth	Zip Code	House number	Street	Place of residence	Country	Linked GGD	Phone number	E-mail	BSN
23098283	Leendert	Van der Plas	01/01/1999	Verify Zip	70	Rapenburg	Leiden	Netherlands	Hollands Midden	0612345678	Leendert@leiden.nl	Verify BSN ....

Table 24 - Database Entries 5 - Verification Functions

Identifier	First Name	Surname	Date of Birth	Zip Code	House number	Street	Place of residence	Country	Linked GGD	Phone number	E-mail	BSN
23098283	Leendert	Van der Plas	01/01/1999	2311 EZ	70	Rapenburg	Leiden	Netherlands	Hollands Midden	Call Number	Send E-mail	123456782

Table 25 - Database Entries 6 - Data replaced by Function

8.5.5 A safe environment to ask and answer questions

Another component of the platform is to provide a safe environment for employees, where they can ask and answer questions in. The reason for this is that it is unlikely that training will ever be effective to the degree that no user will have any questions, especially given the expected large number of employees and the expected lack of time for lengthy training programs. Therefore, organizations must anticipate this need and create an environment where this can be done safely, without leading to additional data breaches such as the ones discussed in Sections 5.4.4 and 5.6.5. This subsection will elaborate on the benefits of this component and provide a practical implementation of this concept, which has been combined with aspects from open-source development and the wisdom of the crowd effect.

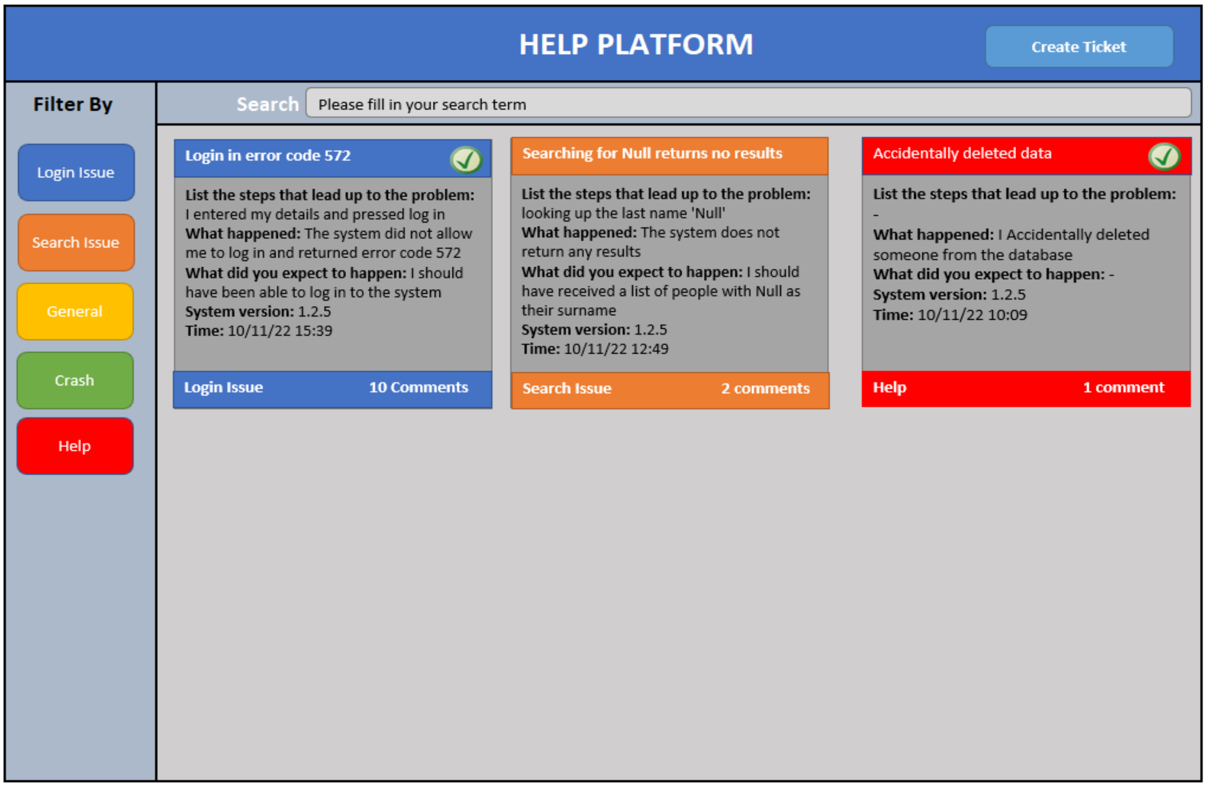
The idea itself also isn't exclusive to a FAIR-based framework, as it could be implemented in any network. However, the data-visiting concept of FAIR may make this a requirement as this is the only way to ensure that data remains contained within the system. Therefore, while it would be a valuable addition to any framework, it may be a requirement for any comprehensive implementation of FAIR.

The benefits of this component are that no data will leave the confines of the system, that appropriate technical measures can be taken in this environment to make it safer than alternatives, and that any attempt to circumvent this environment can be regarded as malicious and be sanctioned accordingly. Additionally, this specific practical implementation adds the benefits of previous knowledge so that users do not have to wait for an answer in some instances and the benefit of collaborating with a large group of employees, by making questions open to all employees and keeping a record of previous questions.

This framework's implementation of this component goes beyond the minimum requirements, taking inspiration from the way tickets are created in technical projects on platforms such as GitHub. This approach has several benefits, including the ability to store and search tickets in a database, the ability to share detailed configuration data, and the ability to automatically track the history of actions taken to resolve issues. The minimal requirements of this component would be to incorporate the environment into the system directly and remove all information that could result in a personal data breach.

When the user encounters a problem or has a question they want to ask, they would make use of a menu button incorporated into the platform, which would open a menu comparable to the menu depicted in Figure 45. The menu includes a search function and filter options, allowing users to investigate whether their problem has already been reported and potentially find a solution. If no solution exists, the user can create their own ticket using a standardized format that automatically includes relevant information such as the system version and time. Other users can also open and comment on existing tickets, providing assistance and contributing to the "wisdom of the crowd" effect. While the exact details of the implementation may vary depending on the specific project, the features included in this example should be considered the minimum requirements for such a system.

The approach discussed above could also be separated into two buttons with one opening a menu and the second directly creating a ticket for their problem, however, that approach would likely result in many unnecessary tickets being created as problems faced by users and the questions they may have are unlikely to be unique. By opening a menu first where the user can search through previous tickets, they would likely be able to find their problem there already answered.



Login in error code 572 ✓

List the steps that lead up to the problem: I entered my details and pressed log in
What happened: The system did not allow me to log in and returned error code 572
What did you expect to happen: I should have been able to log in to the system
System version: 1.2.5
Time: 10/11/22 15:39

Login Issue — 10 Comments

Searching for Null returns no results

List the steps that lead up to the problem: looking up the last name 'Null'
What happened: The system does not return any results
What did you expect to happen: I should have received a list of people with Null as their surname
System version: 1.2.5
Time: 10/11/22 12:49

Search Issue — 2 comments

Accidentally deleted data ✓

List the steps that lead up to the problem: -
What happened: I Accidentally deleted someone from the database
What did you expect to happen: -
System version: 1.2.5
Time: 10/11/22 10:09

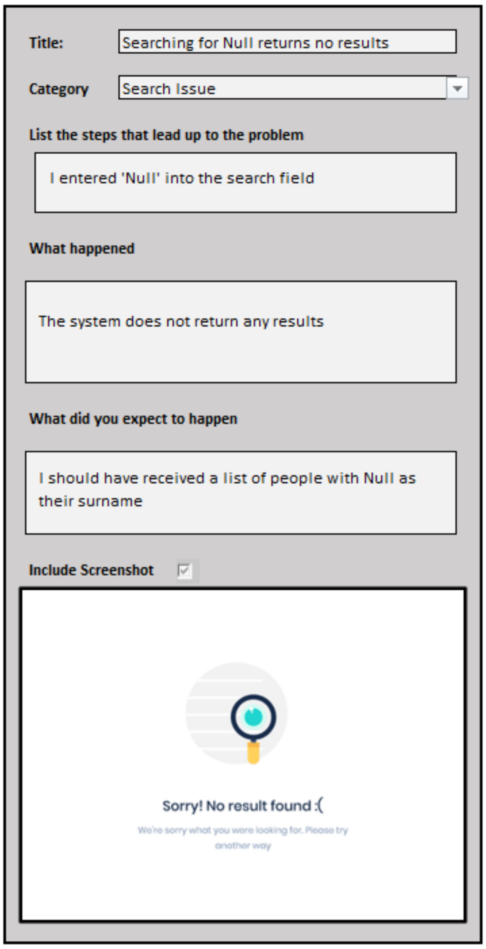
Help — 1 comment

Figure 45 - Example Implementation of the Help Platform

Three example tickets have already been created. The left ticket depicts an issue with logging in to the system. There are 10 comments which indicate that people are either also encountering the issue or that there is an explanation for why the problem occurs. The green checkmark at the top shows that the problem has been resolved. The middle ticket depicts an issue with searching in the system. In this case, there are only 2 comments. Searching for null has caused the query to be thrown out, resulting in no results. This has identified an issue with queries that need to be resolved by a developer. The right ticket shows a person that needs help. In this case, the user has deleted records that need to be recovered.

When clicking on the button 'create ticket' at the top right of the platform, the menu depicted in Figure 46 will open. This opens a menu with a standardized format for reporting issues, allowing users to clearly describe the steps they took, the resulting outcome, and their expected outcome. Screenshots of the software can also be included, with confidential information automatically removed to prevent data breaches.

When clicking on an already created ticket, the menu depicted in Figure 47 will open. This depicts the information from the ticket that has been created, as well as comments that have been placed by either the original poster or other individuals. If the issue appears to be linked to other issues, this information is displayed as well.



Title: Searching for Null returns no results

Category: Search Issue

List the steps that lead up to the problem: I entered 'Null' into the search field

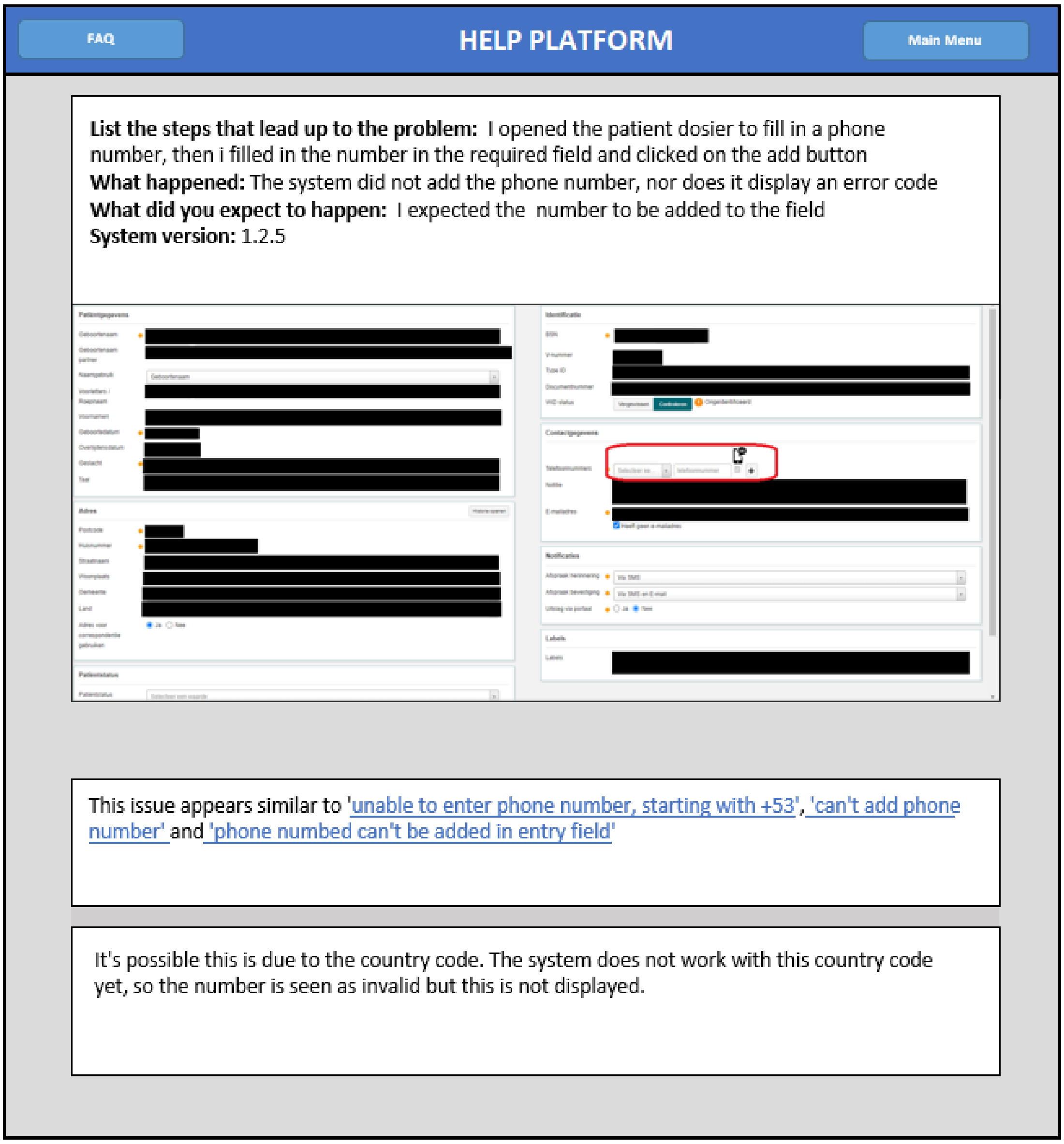
What happened: The system does not return any results

What did you expect to happen: I should have received a list of people with Null as their surname

Include Screenshot ☑

Sorry! No result found :(
We're sorry what you were looking for. Please try another way

Figure 46 – Example of the Ticket Creation Menu



List the steps that lead up to the problem: I opened the patient dosier to fill in a phone number, then i filled in the number in the required field and clicked on the add button
What happened: The system did not add the phone number, nor does it display an error code
What did you expect to happen: I expected the number to be added to the field
System version: 1.2.5

Patiëntgegevens
Geboortenaam [redacted]
Geboortenaam partner [redacted]
Naamgebruik Geboortenaam
Voorletters / Roepnaam [redacted]
Voornamen [redacted]
Geboortedatum [redacted]
Overlijdensdatum [redacted]
Geslacht [redacted]
Taal [redacted]

Adres
Postcode [redacted]
Huisnummer [redacted]
Straatnaam [redacted]
Woonplaats [redacted]
Gemeente [redacted]
Land [redacted]
Adres voor correspondentie gebruiken ● Ja ○ Nee

Patiëntstatus
Patiëntstatus Selecteer een waarde

Identificatie
BSN [redacted]
V-nummer [redacted]
Type ID [redacted]
Documentnummer [redacted]
WID-status Vergoeisen Controleren Ongeïdentificeerd

Contactgegevens
Telefoonnummers Selecteer en... Telefoonnummer
Notitie [redacted]
E-mailadres [redacted]
☑ Heeft geen e-mailadres

Notificaties
Afspraak herinnering Via SMS
Afspraak bevestiging Via SMS en E-mail
Uitslag via portaal ○ Ja ● Nee

Labels
Labels [redacted]

This issue appears similar to 'unable to enter phone number, starting with +53', 'can't add phone number' and 'phone numbed can't be added in entry field'

It's possible this is due to the country code. The system does not work with this country code yet, so the number is seen as invalid but this is not displayed.

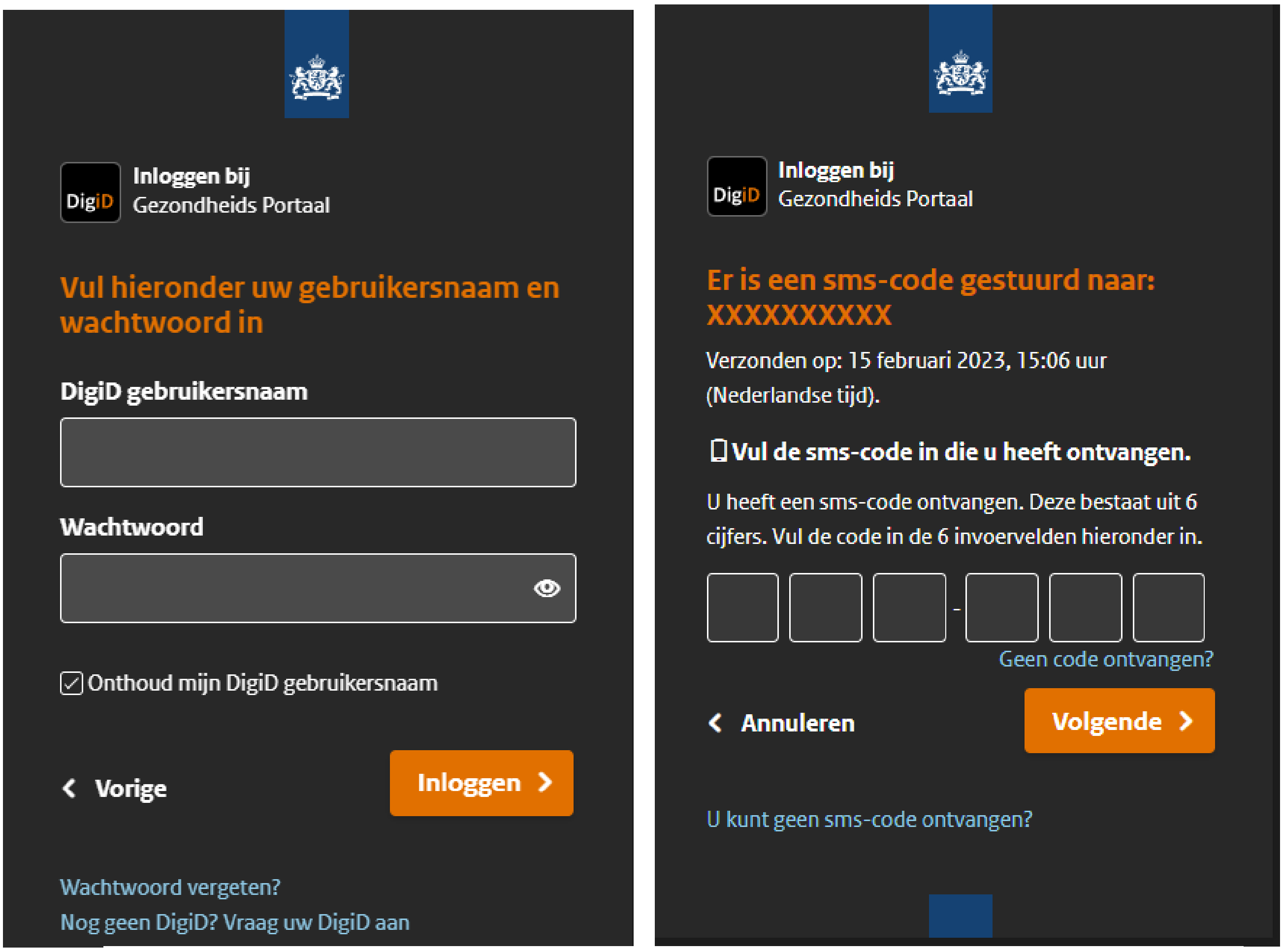
Figure 47 - Example of the Ticket Interaction Screen

8.5.6 Data Subject Access

Another component of the platform is to provide the data subject with read-only access to their data. Using a technological solution, in addition to the normal organizational measures that have to be in place to comply with the rights of the data subject. This subsection will elaborate on how the various rights of the data subjects can be addressed using this approach, in addition to providing a practical implementation of this concept. It should be noted however that this cannot completely replace the traditional way of contacting an organization to make use of these rights as a purely technical solution would likely exclude many older individuals. Which would likely be a violation of the GDPR in its own right.

The idea itself also isn't exclusive to a FAIR-based framework, as it could be implemented in any network. However, to do this successfully does require many of the elements that FAIR addresses to be implemented. For example, to be able to provide the data subject with an overview of all their data, including their usage, requires an organization to have a complete overview of all data usage in their organization, as well as whoever data was shared to. It would also require organizations to make use of a common standard so that data may be erased, updated, or restricted in all locations where data related to the data subject is being processed. Without these requirements, this component would only be able to offer an incomplete overview that would still require significant actions from the organizations to fully comply with the rights of the data subject.

This framework's implementation of this component is based on using the BSN of the data subject as the primary key, which all healthcare systems are required by law under the "Wbsn-z" to register as well as to use whenever data about any patient is exchanged [230] and on using DigID, which is the primary authentication method for government organizations and organizations with a public task such as ministries, local governments, organizations in the healthcare sector, education, pensions, and regional water authorities [231]. Using DigID, the data subject can verify that the BSN belongs to this individual, and by giving access permission to the holder of the BSN by default, secure access can be given to all data belonging to the data subject. This process has been visualized in Figure 48.



Inloggen bij
Gezondheids Portaal

Vul hieronder uw gebruikersnaam en wachtwoord in

DigiD gebruikersnaam

Wachtwoord

☑ Onthoud mijn DigiD gebruikersnaam

Inloggen bij
Gezondheids Portaal

Er is een sms-code gestuurd naar: XXXXXXXXXX

Verzonden op: 15 februari 2023, 15:06 uur (Nederlandse tijd).

Vul de sms-code in die u heeft ontvangen.

U heeft een sms-code ontvangen. Deze bestaat uit 6 cijfers. Vul de code in de 6 invoervelden hieronder in.

Figure 48 - Data Subject Login Process

After the data subject gains access to their data, the data subject can then exercise their data ownership-related rights. A concept mockup of which is visualized in Figure 49. However, this does not mean that no organizational involvement is required anymore. While a user can always be allowed to access their data, the other rights are more complicated and would require an organization's permission in many instances. The platform could aid the right of rectification by allowing the data subject to provide the correct information directly, but an organization could require manual approval for any such changes. The right to erasure and the right to restrict processing could also be initiated by the data subject, but depending on the specific categories, the organization might not want to make this an automatic process. While categories that already have an evaluation that determines that they could be removed or restricted could be handled automatically, in instances where such an evaluation has not taken place or where it was denied, the organization would have to process this manually or process any objections based on a denied decision. The right of portability can be fully automated on a technical level, as FAIR standards are already machine-readable by default, however, organizational approval may still be warranted before initiating any transfer.

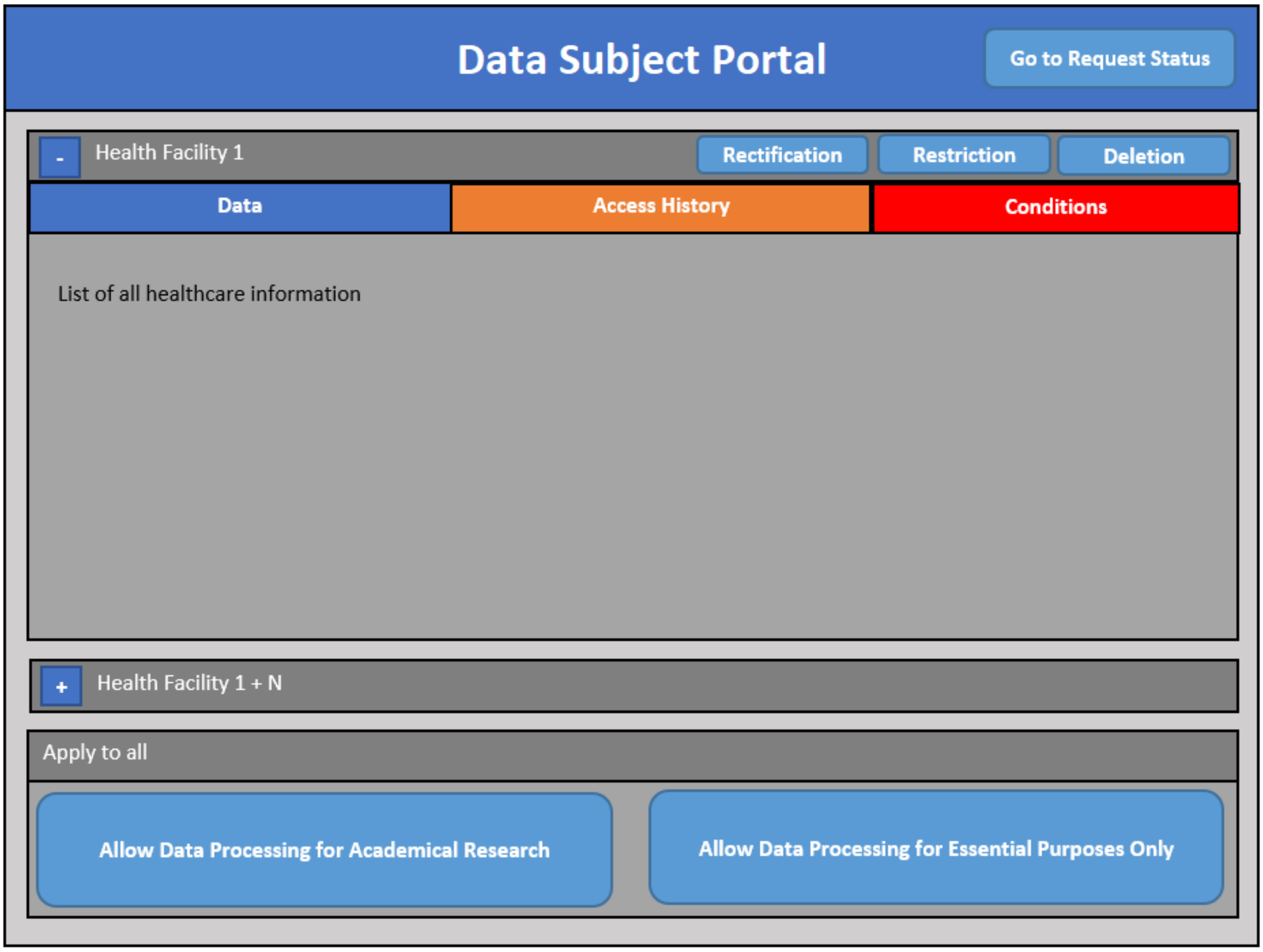


Figure 49 - Concept Mockup of the Data Subject Portal

This component is also able to exceed the requirements set out by the GDPR, giving an even higher level of control to the data subject. Data subjects could be given the option to decide for what purposes their data could be re-used, for example in the context of research. Using an opt-in or opt-out approach could make significantly more data accessible for research in a user-friendly way, however, the legality of such an opt-out approach would have to be investigated and an opt-in approach would have to be actively encouraged to make as many people as possible provide access to their data. It should also be noted that the GDPR technically does not require this, as Article 87 of the GDPR provides exceptions where data subjects do not have to give consent for the processing of their data. However, as organizations are hesitant to provide access to their data when consent is not given by the data subject, even with the existence of Article 87, this approach would eliminate this reasoning.

8.6 Governance Process

The proposed FAIR-based framework can be applied to project development in three primary circumstances, resulting in three distinct approaches: creating a project based on the default version of the framework, developing a project using a similar project as a starting point, and deploying another instance of a project when minimal or no changes are needed. The central registry can be utilized to determine which approach to select by identifying any existing similar projects that match the desired specifications and requirements, as illustrated in Figure 50. Regardless of the chosen approach, some level of facility-specific changes and additional configuration will be necessary, with approaches 2 and 3 reducing or eliminating other required modifications.

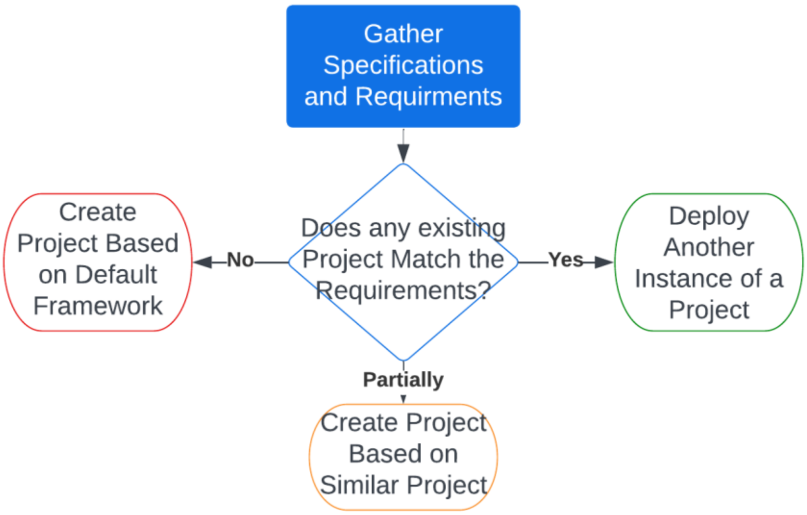

Gather Specifications and Requirments
Does any existing Project Match the Requirements?
No
Create Project Based on Default Framework
Yes
Deploy Another Instance of a Project
Partially
Create Project Based on Similar Project

Figure 50 - Selecting Project Development Approach

The first approach is employed when no similar projects based on the common framework exist, requiring developers to start from the default version of the framework. As the framework's usage expands, this approach will likely become less frequent, with more similar projects serving as better starting points for new projects. Although the default version contains all components described in this thesis, it lacks functionalities tailored for specific projects. This approach is discussed in more detail in Section 8.6.1.

The second approach is employed when existing projects provide a better foundation for new project development. As projects are rarely unique and can benefit from previous implementations, this approach will likely become the most common method for developing projects using the proposed FAIR-based framework. By reducing the workload needed to achieve a certain level of functionality, the second approach streamlines the development process. For example, starting from another implementation of a source and contact tracing investigation system would be more efficient than using the default version of the framework. This approach is discussed in more detail in Section 8.6.2.

The third approach is implemented when previously created projects can be deployed in another location without extensive changes. This method enables new facilities to quickly adopt a proven, safe, and secure system. It is particularly useful for similar organizations with identical tasks, such as hospitals, clinics, and GGDs. For instance, any hospital could use the same software since they perform the same tasks, allowing for swift deployment and full data control. In crisis situations, this approach facilitates rapid capability increases when needed, such as creating a federated version of CoronIT with each facility only accessing its own data. This approach is discussed in more detail in Section 8.6.3.

Section 8.6.4 discusses how data from systems not built using the proposed FAIR-based Framework could be imported into a system that was built based on the proposed FAIR-based Framework. While this is not the intended use case of this Framework, there are two situations that would make this likely, either when migrating from a previous system or in a wrongful application of the Framework.

8.6.1 Creating a Project Based on the Default Framework

The first step in each approach is to deploy an iteration of the proposed FAIR-based framework, generating the necessary components, data files, and organizational documentation to both comply with GDPR and create an efficient, effective system. This can be achieved using a git or container-based solution. Initially, the framework applied consists only of its core components, lacking implementation of functionalities and metadata descriptions of FAIR data, as no data is included in the default framework. A DPIA template, based on the technology related to the framework, is provided but incomplete due to the absence of data processing information and processing rationale. Figure 51 illustrates the deployed core framework, while Figure 52 provides an overview of all steps.

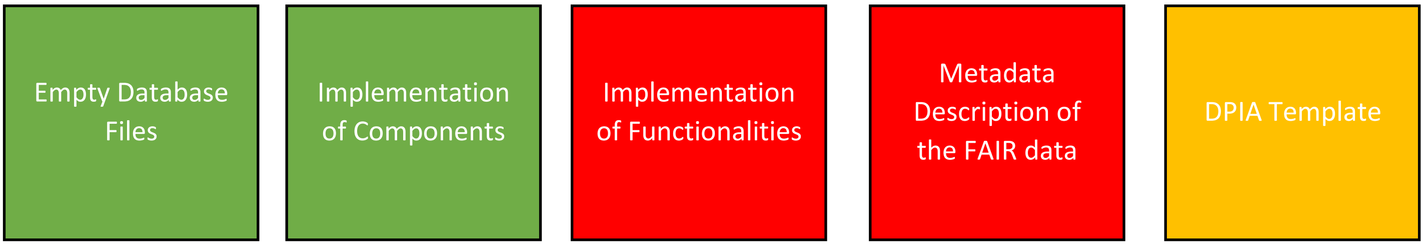

Empty Database Files
Implementation of Components
Implementation of Functionalities
Metadata Description of the FAIR data
DPIA Template

Figure 51 - Completion status after deployment - Approach 1

To promote transparency and align with the government's open-source objectives, software development should occur in a public development branch. A version should only be deployed and made accessible to employees upon receiving approval and a positive DPIA report. Critical security updates are exempt from this process but should never introduce new functionalities, to prevent his process from acting as a workaround to the approval process. The open-source nature of the FAIR-based framework minimizes the likelihood of any such attempt going undetected.

After successful deployment, facility-specific metadata must be added to customize the project for the particular facility. Section 8.1 details this information, such as facility name, system purpose, operation, and data storage duration.

The development process follows an Agile methodology, using iterative cycles to implement specifications and requirements, gather feedback, and make improvements. Alternative methodologies are possible but should be suitable for rapid development and feedback processing. For example, the waterfall methodology's linear progression may hinder development speed and feedback integration, especially during a crisis, and therefore be an ineffective methodology for us in combination with the proposed FAIR-based framework.

Additionally, the common framework may prove to be a benefit even if another project has not been chosen to act as the starting point. Any project can benefit from the specific implementation of individual system components and functionalities that have been created for another system based on the common framework, by more easily incorporating these elements in a new system. This saves development costs, and time. Additionally, any already implemented feature is less likely to result in a security vulnerability given that any released version would have to be in full compliance with the GDPR.

Upon sufficient development progress, a new or updated DPIA can be created using the template DPIA for the default framework. If a DPIA has already been created, a newer version can be created by updating the previous one. This process can further be improved if features are incorporated from previously existing systems, by incorporating the corresponding section of the DPIA from that system. In the event that the DPIA contains any significant risk that cannot be mitigated, the DPIA must be reported to the AP for them to evaluate the DPIA and either allow or bar the current version of the system from being released. This framework has incorporated this legal requirement, yet given the importance of any such system, it may be warranted to make the submission of the DPIA a requirement for any project, no matter the presence of any significant risks. While there would be no legal obligation to wait for a response from the AP in these instances, the AP would be able to provide advice on any of the content or may even prevent an unsafe project from continuing.

If the AP delivers a negative judgment, the project returns to the development cycle to address their concerns. Once adequately addressed, the DPIA must be updated and reviewed by the AP again.

Upon release approval, changes are made at both the central registry and the corresponding facility. The approved version is deployed at the facility and made available to users. If this is an entirely new project, the system must first be made accessible to clients by registering the internal ports and local IP address to a publicly accessible address and uploading this information to the central registry. After user creation and role assignment, users can access the system with their credentials and perform their duties and responsibilities. At the central registry, the DPIA is uploaded and made accessible directly at the projects page on the registry, in addition to all other relevant information that has been previously created.

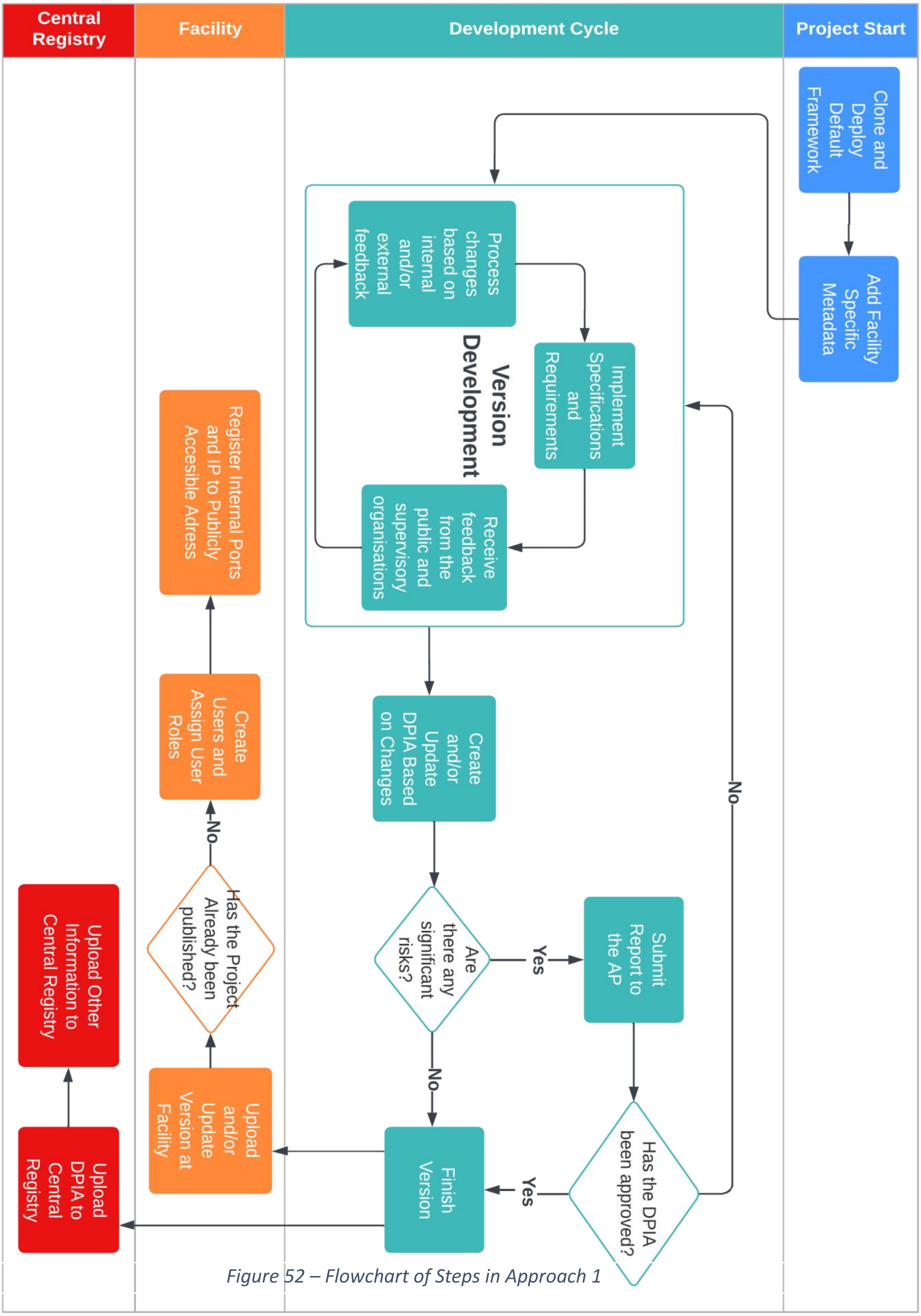

Project Start
Development Cycle
Facility
Central Registry
Clone and Deploy Default Framework
Add Facility Specific Metadata
Version Development
Process changes based on internal and/or external feedback
Implement Specifications and Requirements
Receive feedback from the public and supervisory organisations
Create and/or Update DPIA Based on Changes
Are there any significant risks?
Yes
Submit Report to the AP
Has the DPIA been approved?
No
Yes
No
Finish Version
Upload and/or Update Version at Facility
Has the Project Already been published?
No
Create Users and Assign User Roles
Register Internal Ports and IP to Publicly Accesible Adress
Upload DPIA to Central Registry
Upload Other Information to Central Registry

Figure 52 – Flowchart of Steps in Approach 1

8.6.2 Creating a Project Based on a Similar Project

The first step in each approach is to deploy an iteration of the proposed FAIR-based framework, generating the necessary components, data files, and organizational documentation to both comply with GDPR and create an efficient, effective system. This can be achieved using a git or container-based solution. In this approach, the version of the framework that is deployed is based on a project sharing similarities with the to-be-developed project, offering a more developed starting point. Figure 53 illustrates the deployed similarity-based framework, while Figure 54 provides an overview of all steps.

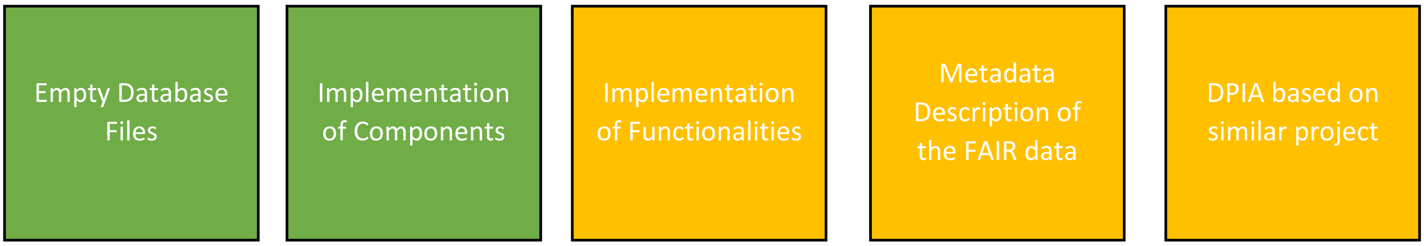

Empty Database Files
Implementation of Components
Implementation of Functionalities
Metadata Description of the FAIR data
DPIA based on similar project

Figure 53 - Completion status after deployment - Approach 2

To promote transparency and align with the government's open-source objectives, software development should occur in a public development branch. A version should only be deployed and made accessible to employees upon receiving approval and a positive DPIA report. Critical security updates are exempt from this process but should never introduce new functionalities, to prevent his process from acting as a workaround to the approval process. The open-source nature of the FAIR-based framework minimizes the likelihood of any such attempt going undetected.

While this approach also requires a development cycle, this cycle is likely to be less substantial than the previous approach which would only deploy the essentials of the framework. Development work has already been done on the implementation of functionalities, the metadata description of the data contained in the system, and a DPIA containing this information. The to-be-developed project can then build on this project, adding additional features according to the specifications and requirements of the system. It should however be noted that this approach needs to carefully examine legacy functionalities as any of these functionalities are likely to be created based on different situations and different requirements. The to-be-developed project needs to either remove such features if they are not required or keep them included in the DPIA, as these functionalities can potentially become a vulnerability or security risk.

The development process of this second approach is similar in the steps that are followed compared to the first approach, however, this is intended. The main difference between the two approaches is the amount of work that is required to create a system that meets the specifications and requirements of the project. Using this approach, the amount of work should be substantially lower due to the possibility of re-using previously implemented functionalities. This becomes more effective as the framework matures and more projects would exist for any circumstance, resulting in more similar projects existing that can serve as an even better starting point. Additionally, the same concept of borrowing implementations from all other projects can also be applied to the second approach, resulting in an even more efficient development process.

After successful deployment, facility-specific metadata must be added to customize the project for the particular facility. Section 8.1 details this information, such as facility name, system purpose, operation, and data storage duration.

The development process follows an Agile methodology, using iterative cycles to implement specifications and requirements, gather feedback, and make improvements. Alternative methodologies are possible but should be suitable for rapid development and feedback processing. For example, the waterfall methodology's linear progression may hinder development speed and feedback integration, especially during a crisis, and therefore be an ineffective methodology for us in combination with the proposed FAIR-based framework.

Upon sufficient development progress, a new or updated DPIA can be created using the template DPIA for the default framework. If a DPIA has already been created, a newer version can be created by updating the previous one. This process can further be improved if features are incorporated from previously existing systems, by incorporating the corresponding section of the DPIA from that system. In the event that the DPIA contains any significant risk that cannot be mitigated, the DPIA must be reported to the AP for them to evaluate the DPIA and either allow or bar the current version of the system from being released. This framework has incorporated this legal requirement, yet given the importance of any such system, it may be warranted to make the submission of the DPIA a requirement for any project, no matter the presence of any significant risks. While there would be no legal obligation to wait for a response from the AP in these instances, the AP would be able to provide advice on any of the content or may even prevent an unsafe project from continuing.

If the AP delivers a negative judgment, the project returns to the development cycle to address their concerns. Once adequately addressed, the DPIA must be updated and reviewed by the AP again.

Upon release approval, changes are made at both the central registry and the corresponding facility. The approved version is deployed at the facility and made available to users. If this is an entirely new project, the system must first be made accessible to clients by registering the internal ports and local IP address to a publicly accessible address and uploading this information to the central registry. After user creation and role assignment, users can access the system with their credentials and perform their duties and responsibilities. At the central registry, the DPIA is uploaded and made accessible directly at the projects page on the registry, in addition to all other relevant information that has been previously created.

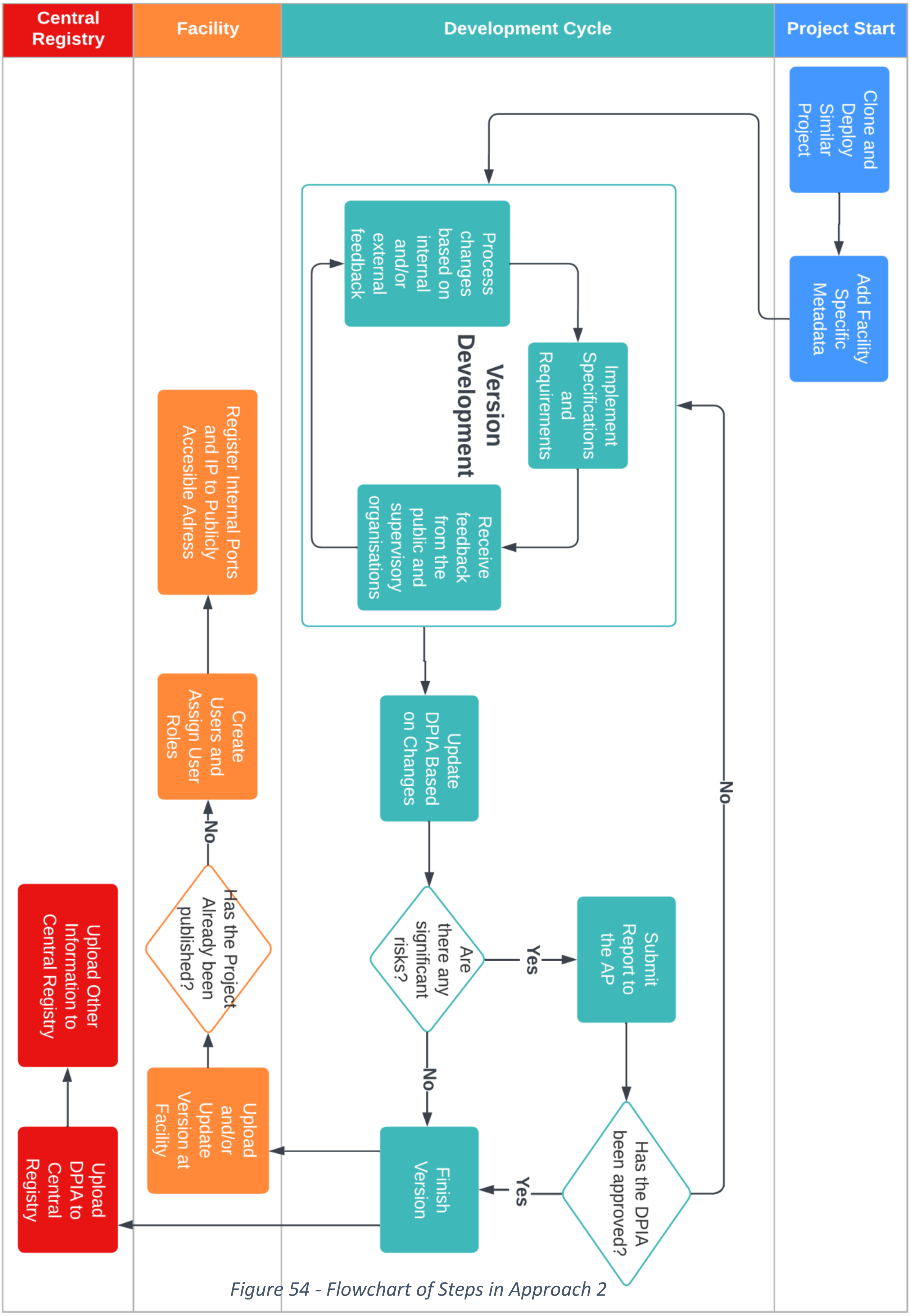

Central Registry
Facility
Development Cycle
Project Start
Clone and Deploy Similar Project
Add Facility Specific Metadata
Version Development
Process changes based on internal and/or external feedback
Implement Specifications and Requirements
Receive feedback from the public and supervisory organisations
Update DPIA Based on Changes
Are there any significant risks?
Yes
No
Submit Report to the AP
Has the DPIA been approved?
Yes
No
Finish Version
Upload and/or Update Version at Facility
Has the Project Already been published?
No
Create Users and Assign User Roles
Register Internal Ports and IP to Publicly Accesible Adress
Upload DPIA to Central Registry
Upload Other Information to Central Registry

Figure 54 - Flowchart of Steps in Approach 2

8.6.3 Deploying Another Instance of a project

The first step in each approach is to deploy an iteration of the proposed FAIR-based framework, generating the necessary components, data files, and organizational documentation to both comply with GDPR and create an efficient, effective system. This can be achieved using a git or container-based solution. In this approach, the version of the framework that can be applied requires only minor changes, which will be specific to the current facility. An overview of what has been deployed can be seen in Figure 55, which depicts that all aspects of the system are either fully or nearly complete immediately after deployment. An overview of all steps has been depicted in Figure 56.

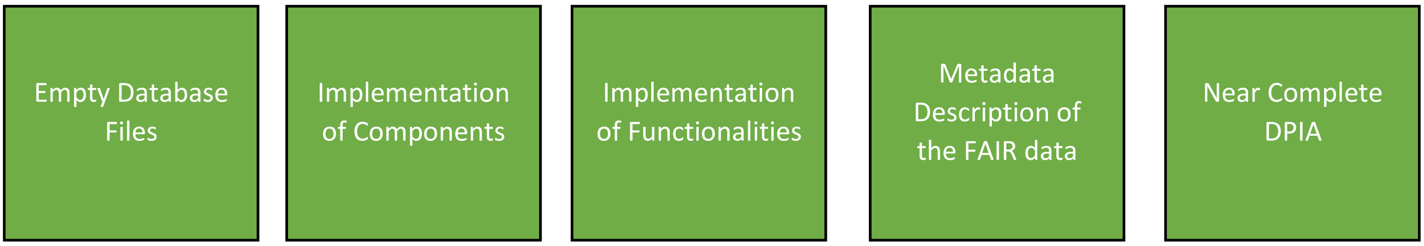

Empty Database Files
Implementation of Components
Implementation of Functionalities
Metadata Description of the FAIR data
Near Complete DPIA

Figure 55 - Completion status after deployment - Approach 3

After successful deployment, facility-specific metadata must be added to customize the project for the particular facility. Section 8.1 details this information, such as facility name, system purpose, operation, and data storage duration. Unrelated to the data files, this requires setting up protocols for access to these files as well as protocols related to the monitoring system such as who is responsible for manually investigating users that have been flagged or blocked for their activity.

The next step is to update the nearly complete DPIA with protocols related to who can access sensitive files and the monitoring system. If the DPIA of the previous project has been determined to be safe enough to be allowed to process data, the DPIA of this project would likely be determined to be safe enough to allow for the processing of data as well. Resulting in a GDPR-compliant system that can be released.

If the AP delivers a negative judgment, given previous approval for the project this project has been cloned from, the most likely reasoning is an incorrect implementation of sensitive protocols. These protocols must therefore be adjusted to ensure that they are in line with the GDPR. Once adequately addressed, the DPIA must be updated and reviewed by the AP again.

Upon release approval, changes are made at both the central registry and the corresponding facility. The approved version is deployed at the facility and made available to users. If this is an entirely new project, the system must first be made accessible to clients by registering the internal ports and local IP address to a publicly accessible address and uploading this information to the central registry. After user creation and role assignment, users can access the system with their credentials and perform their duties and responsibilities. At the central registry, the DPIA is uploaded and made accessible directly at the projects page on the registry, in addition to all other relevant information that has been previously created.

Considering that this approach entails utilizing a pre-existing system with minor, facility-specific modifications, it is optimal for all facilities to collaboratively develop updates related to security or functionality. This is particularly crucial for security updates that tackle vulnerabilities affecting all facilities. Although one facility may implement functionality updates that could benefit others, this is likely an exception rather than the norm. The collective resources and capabilities of all facilities utilizing the same system significantly surpass those of an individual facility.

Updates can be seamlessly applied through the same system used for the initial deployment, replacing existing data with verified information from a trusted source. Validation methods, such as checksums, ensure that the downloaded data matches the intended information from the trusted source. Importantly, this process should not overwrite any facility-specific modifications.

A key advantage of this approach is the ability to rapidly scale capacity without compromising privacy and security. However, it cannot guarantee absolute privacy and security since this relies on more than just technology. Although the necessary organizational measures are integrated into this framework, they cannot ensure their proper execution. For instance, establishing protocols for accessing the central files of the project poses a significant security risk; if compromised, the entire facility's database would be endangered. Additionally, no technical solution can address the issue of granting excessive access to numerous users. Therefore, while this approach streamlines and secures the process, it does not guarantee security, emphasizing the importance of properly executing each step.

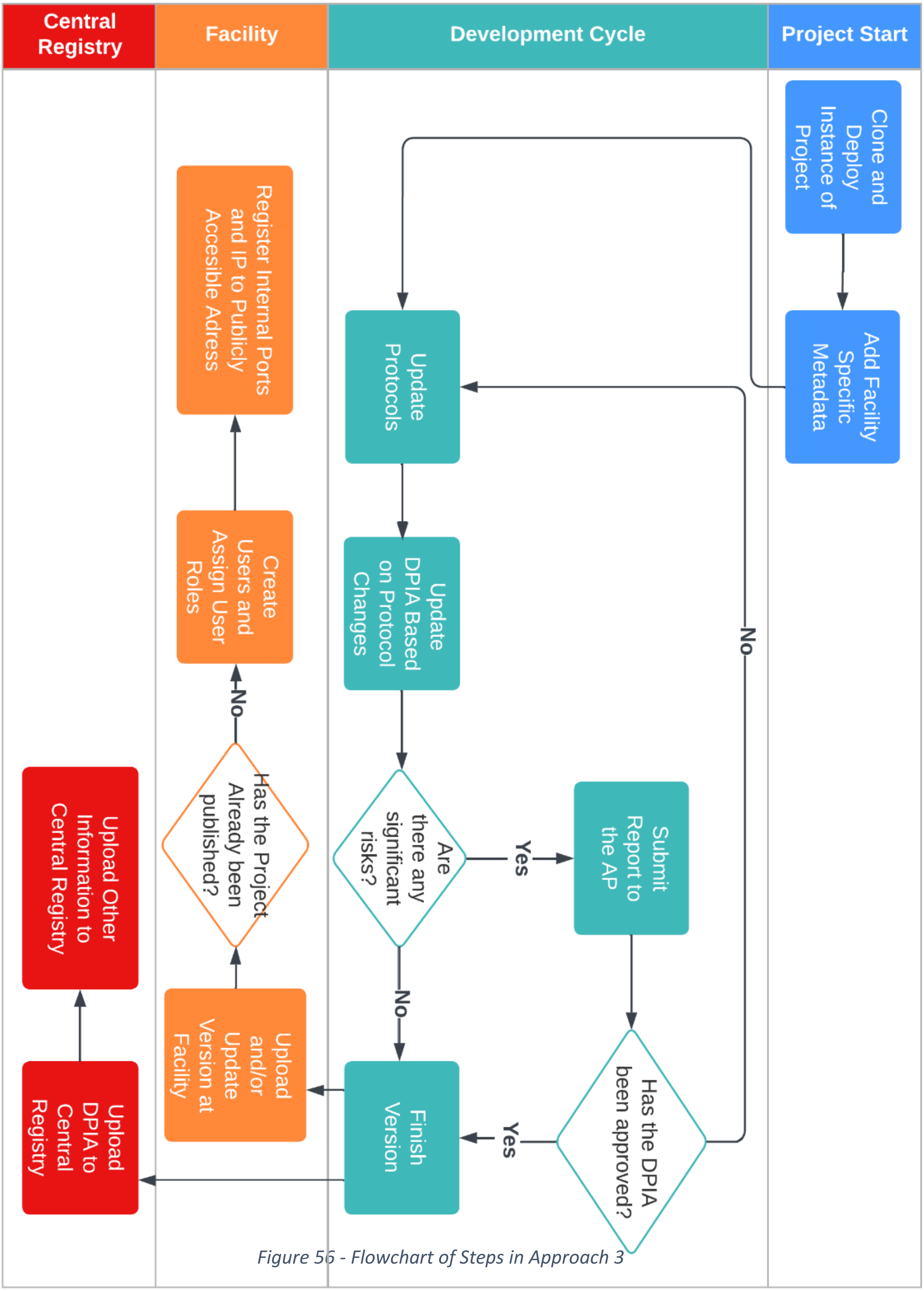

Project Start
Development Cycle
Facility
Central Registry
Clone and Deploy Instance of Project
Add Facility Specific Metadata
Update Protocols
Update DPIA Based on Protocol Changes
Are there any significant risks?
Yes
Submit Report to the AP
Has the DPIA been approved?
No
Yes
No
Finish Version
Register Internal Ports and IP to Publicly Accesible Adress
Create Users and Assign User Roles
No
Has the Project Already been published?
Upload and/or Update Version at Facility
Upload Other Information to Central Registry
Upload DPIA to Central Registry

Figure 56 - Flowchart of Steps in Approach 3

8.6.4 Importing Data into a FAIR-based Project

Given the prevalence of legacy systems that currently have not been required to meet modern regulatory and security standards, there is a significant number of systems that will need to be replaced by more modern counterparts. More modern systems may also need to be replaced in the event that they are unable to meet these same standards. To promote the development of these projects using the proposed FAIR-based framework, it is important to make this transition as easy as possible, using the existing data that was already stored in these previous systems. This section will describe this process in more detail.

The first step of this process would be to develop the project based on one of the three previous development methods. Deploying the default framework if no similar projects have previously been created, or deploying a previously created project or even a different instance if a similar project has been created before. After this step is completed, a fully completed system without any data will have been created.

After completion of the system, data needs to be added to the system. However, given that the data from the previous system has not been FAIR-ified it is currently unable to be directly imported. Not only is it unlikely that the data variables will share a common terminology and/or format with the new project, but it also lacks any relations between data that are part of the ontology. To ensure that data remains interchangeable, the data must first be processed and transformed to fit into this ontology.

Therefore, the second step of this process is to prepare the data for use in the ontology. To do this, all data columns need to be examined to determine if they represent entities, attributes, or relationships in the ontology. Data columns that represent ontology elements can then be mapped to their corresponding elements. Data may also require additional processing, for example, if the information is stored in a different format than what is specified in the ontology. For example, dates may be stored in a string format in the original system, while the governmental ontology requires this information to be stored in the form of a "DateTime".

However, even with this process, some data may remain incompatible with the ontology in which this data should not be included as it would lead to inconsistent values. For example, the previous data set may have stored the fever attribute, with no reference to what constitutes a fewer. The ontology meanwhile uses a strict definition with a fever being defined as a body temperature exceeding a certain body temperature. In this instance, the information could be included using a different relation, to temporarily make use of it in the primary healthcare process, but should not be used in any kind of analysis.

The third and last step of this process is to create RDF triples based on the mapping of the data columns with the corresponding ontology elements. The creation of the triples itself can be achieved using a simple script that reads the format the previous data has been stored in, applies all relevant transformations, and then creates the RDF triple itself. The triples can then be imported into an ontology file, after which the ontology data can be validated against the ontology's restraints concerning consistency, completeness, and correctness. Any mistakes in this process need to be corrected in the previous step. Upon completion of this process, all relevant information has been imported into the new, FAIR-based, system.

8.7 Security Vulnerabilities

This section discusses the security vulnerabilities inherent to the proposed FAIR-based framework, which are not necessarily unique to the proposed FAIR-based framework but are instead vulnerabilities that cannot be fully eliminated. These vulnerabilities can be divided into two categories: security vulnerabilities related to network intrusion and vulnerabilities related to incorrect implementation of the proposed FAIR-based framework.

The examples provided are not an exhaustive list of possible attacks, and should not be considered the only types of attacks an organization must address. Instead, the aim is to raise awareness that, while this architecture is designed to ensure privacy and security, no system can guarantee it. Especially given the concept of zero-day exploits, which are unknown flaws or vulnerabilities, that in some instances cannot even be anticipated.

Continuous work is required to ensure the sensitive personal data of numerous individuals remains secure. The framework's intention to establish a common standard with transparent development significantly increases resources dedicated to addressing these security vulnerabilities and expands the number of eyes on the project to aid in detecting vulnerabilities. However, even this is not a guarantee of security. Project-specific additions and changes also need careful review to ensure they do not result in additional security vulnerabilities.

8.7.1 Network intrusions

Network intrusions pose a major threat to any computer system. Like traditional systems, the proposed FAIR-based framework must protect its data from external attacks. Although this thesis has not extensively covered protection from external attacks, it will discuss the possible effects of a network intrusion by describing the potential impact on specific components and how these can be mitigated. This section will not cover preventing network intrusion, as this is unrelated to the framework itself and depends on organizational security measures or potential flaws in system protection.

A significant network intrusion could result in the database being transferred to a malicious actor. While this would constitute a major personal data breach in scope, with large-scale breaches affecting millions of people, any database should be strongly encrypted, reducing the actual impact on affected individuals. Encryption does not guarantee data privacy and security, though; exposure of the decryption key or flaws in database implementation could still allow malicious actors to recover valuable information. Thus, the focus should be on preventing malicious actors from accessing the database, preventing database transfer, and monitoring and shutting down any transfer attempts quickly.

A less significant network intrusion, or one where the malicious actor does not expect to recover information from the encrypted database, might target metadata files regulating data access. By altering these files, an attacker could adjust the level of information exposed to specific users and user roles. For example, an attacker might modify the search query system to expose complete information on all data subjects without limiting the information returned. Although such an attempt would be noticeable to other users with that role, making it a highly visible attack, an attacker might create a new role and assign themselves to it to avoid detection.

However, these attacks can be prevented by making metadata files immutable, except under very specific and secure circumstances. Any changes in metadata files should be logged with high-security flags or time-based, so any attack can be discovered before a malicious actor can make or exploit these changes. The framework itself could also be designed to disallow such extreme changes, although a malicious actor might then opt for more subtle changes. If the previous system is implemented correctly, even minor changes would be immediately apparent and unlikely to succeed.

The most sophisticated attack would target metadata files only when there is a reason for them to be changed. For example, if a framework is constructed and a specific project is created or updated at a location from a central repository, a malicious actor might target this time to weaken the system's security. Instead of using the central repository, the project could be based on a forked repository with hard-to-detect but significant changes. However, even these kinds of attacks can be detected by employing a hash validation function to ensure updates use the correct project files.

The least significant network intrusion, or ones where the malicious actor aims to remain entirely undetected, involves adding themselves to the user database. By doing this, they would gain access to the same information available to users of a specific role. This type of attack is relatively challenging to detect, although a system could be designed to identify anomalous additions to the user database.

A more subtle and less detectable method would involve using social engineering to add users to the database by exploiting the credentials of an authorized individual. This should be prevented by the multi-factor authentication system required for account access or by a monitoring system that identifies such attempts as aberrant behavior. Furthermore, the monitoring system should detect unusual behavior when the malicious actor logs in as a legitimate user and tries to extract information that deviates from the activities of regular employees.

8.7.2 Incorrect implementation of the FAIR-based framework

In addition to network intrusions, security vulnerabilities may also arise from mistakes made during the creation of the proposed FAIR-based Framework or from how an organization configures the system developed using the framework.

To give an example of the first category, the entire system around data access is only able to guarantee the privacy and security of the data subject if the system itself doesn’t contain any exploitable flaw. While this could be exploited through a network intrusion, there is also a major danger from internal employees that already have legitimate access to the system. Through exploiting vulnerabilities, they would be able to access information beyond the restrictions of their user account and user role.

To prevent this, every API call, a user's role, identity, and level of access must be validated at every step. The various flowcharts and pseudocode included in Sections 8.3.1 to 8.3.3 offer an example of this with it never being assumed that a user has the required level of access simply because the user was able to call that function. However, there are many exploits related to bypassing lines of code execution such as a buffer flow for example, which could prevent these checks from being executed. Meanwhile, there are also various techniques to prevent these exploits, such as Address space layout randomization to counter the specific previous example. To try to guarantee the privacy and security of the data subject, this will need to be done for everyone's known exploit, as well as anticipation of future possible exploits. While this would likely be beyond the level of resources committed to a single project, this is entirely possible to commit to the further development of the common framework, as this proposed FAIR-based framework has been designed to be.

To give an example of the second category, while the access and control system can effectively restrict data and information access to any user and thereby guarantee the privacy and security of the data subject, the highly flexible nature of this system allows organizations to determine themselves what these restrictions are. In the event that organizations aren't restricitive enough in providing access, users could still end up in situations where they have access to far more data than they would require. This would result in a repeat of what happened to CoronIT, even with a system that is superior concerning the protection of privacy and security in every way. Given that these problems are organizational instead of technical, a technical solution cannot address this without hampering the flexible nature of the system.

The proposed FAIR-based framework contains some elements to prevent this from happening, however, the effectiveness of these techniques is uncertain. For example, a DPIA would state the risks contained in the system, but as this is the way that systems traditionally operate, Data Protection Officers may not see these decisions as a risk. With the system containing no significant risks, It is not legally required to report the DPIA to the AP. While the proposed FAIR-based framework suggests making this a mandatory step, this can only be enforced by the Dutch government. AcICT supervision would theoretically be able to spot such decisions, however, given both the budgetary restrictions of the organization and their mandate is limited to larger projects, not all projects would be able to be reviewed.

The open-source nature of the FAIR-based Framework and the increased level of transparency through the use of the central registry may aid in such decisions being detected, but from an outside point of view, these decisions aren't necessarily wrong and depend on the expertise of the staff involved in processing the data and their requirements and responsibilities. It is likely that in many circumstances, these things would go unnoticed. Especially at the start of the use of the common framework where nobody has experience with either its potential or normal mode of operation. However, such information could be included from the start in the central registry.

Another possibility would be to limit the flexibility or create some sort of warning system based on previous knowledge with specific data fields and the number of employees and level of access that is provided. This is a realistic solution but needs additional development to determine the exact details of its implementation, as well as the decisions to allow development but provide a warning, or to prevent a project from being released. Additional considerations need to be made concerning the fact that users are added at a later time, which would change the number of users with access and therefore the potential risks of any level of data access.

Another example of the second category is an incorrect implementation of the monitoring procedure. This risk is not related to the organization's decision to implement a monitoring procedure, nor the fact that such a procedure could be inadequate. While these are possibilities, this thesis will assume that these problems would be spotted through the creation of the DPIA, either by the public, the AP, or the Data Protection Officers involved in the project. Instead, the main risk is in the organization not carrying out its created procedure. No supervisory organization or component guarantees that an organization carries out this process correctly.

A possible solution to address this could be to have organizations report the number of incidents they have, including the number of incidents that required manual review and what the outcome of these reviews was. This would measure the effectiveness of the organization's procedures regarding the monitoring system. Although with the major caveat that if the monitoring system itself would be unable to distinguish behavior, or only able to catch the worst offenders, the results would be positive.

9. Discussion

The discussion section has been divided into four sections. Section 9.1 discusses the case study into the GGD during the Covid-19 pandemic, in the scope of HPZone (Lite), CoronIT, and the further context of the governmental IT environment in the Netherlands. Section 9.2 discusses the reasoning behind the choice to make the proposed FAIR-based Framework a common framework instead of a purely crisis-specific one. Section 9.3 discusses the use case and considerations that this thesis has found for the proposed FAIR-based framework

9.1 Case Study into the GGD during the Covid-19 Pandemic

This case study investigated the role of the GGD during the Covid-19 pandemic, primarily focusing on the main systems used to support source and contact tracing processes (HPZone and HPZone Lite) and testing and vaccination processes (CoronIT). In addition to examining these systems, the study explored alternative options and the broader Dutch IT environment to contextualize the findings. This comprehensive investigation covered potential strategies for system development, including utilizing or improving legacy software (HPZone and HPZone Lite), acquiring commercial off-the-shelf software (CoronIT), and developing software internally (Corona Melder App).

The primary focus was on CoronIT and HPZone (Lite), as they were significant support tools requiring extensive human interaction. These systems were used by tens of thousands of individuals and needed to meet privacy, security, and internal security standards to protect against malicious actors with system access. In contrast, the Corona Melder App was a phone app with less human involvement, reducing internal security risks. However, privacy and security by design remained crucial. Since human presence was the most significant challenge during the pandemic, the Corona Melder App was only used for comparison in certain aspects and was not investigated in detail.

Crisis situations are unique and difficult to anticipate or prepare for, and often require a rapid response. This results in a complex question, where on the one hand a quick response is needed which would imply that any solution needs to either be quickly developed or developed in advance, while on the other hand, the uniqueness of a crisis situation makes it difficult to impossible to either have a solution on hand or to develop one quickly. This case study addresses the global pandemic crisis, with systems supporting source and contact tracing investigation and testing and vaccination processes as the solution. Other crises, whether in healthcare or elsewhere, may pose unique challenges with distinct solutions. Consequently, while the findings apply to this crisis, they may not entirely apply to other situations, although many insights are broadly applicable to other organizations and projects.

In theory, the Dutch government should have been well-prepared for this crisis, as the necessary infrastructure and experienced personnel were in place. The HPZone system, used for years for source and contact tracing investigations, was employed during the pandemic, albeit on a larger scale and with a revised version (HPZone Lite) sharing the same database and architecture. For testing and vaccination processes, Praeventis, the system supporting the Dutch national vaccination program since 2003, would have been used. However, the government decided to use CoronIT, a newly acquired commercial system, instead. As a result, the Dutch government could not fully benefit from the existing systems that would otherwise have been utilized to address the crisis.

HPZone was not designed to support the scale it was used at during the pandemic, with tens of thousands of employees instead of the usual hundreds. The system also failed to meet modern requirements before the pandemic, raising questions about its continued use. The GGD and the Dutch government were aware of these issues but chose to use the system due to a lack of alternatives and a desire to maintain continuity. This contrasts with the decision to replace Praeventis with newly developed software.

A detailed examination of HPZone Lite and CoronIT revealed that despite differing initial approaches to data management, design decisions led to both systems sharing similar flaws and vulnerabilities. Initially, HPZone was a highly federated system with each GGD region managing its instance. Activity by users was generally not monitored and any monitoring logs were only kept for a short duration. During the Covid-19 pandemic, HPZone Lite was created without addressing existing issues but separating Covid-19 information from other diseases and allowing employees from one region to operate in another's instance. This resulted in a system where employees could access information on data subjects across the Netherlands, with unlimited access levels and minimal monitoring systems.

CoronIT, specifically acquired for the crisis, was designed as a unified system allowing employees to access information related to data subjects regardless of their security region. While technical information related to CoronIT is limited, it can be expected that due to being acquired specifically for a major crisis, it was built with respect to the high number of employees that would be interacting with it. Which is supported by the limited technical disruptions related to this system. While being designed for a high number of likely less experienced employees, every employee would still have a high level of access, being able to fill in for any job in any region. Monitoring systems were referenced in the DPIA, and deemed to be required due to the substantial value of the information contained in this system, but were never properly implemented. Resulting in a second system where employees would be able to access information from data subjects across the entirety of the Netherlands, with an unlimited level of access and few to no monitoring systems in place.

Both HPZone and CoronIT systems show substantial evidence of non-compliance with legislation. Section 5.6.7 reveals that the procedures for creating and updating Data Protection Impact Assessments (DPIAs) were not adhered to. Although HPZone's legacy status might justify the absence of a DPIA, the significant changes introduced by the derivative system, HPZone Lite, certainly necessitate the creation of one. CoronIT, as a newly developed or acquired system, always requires a DPIA. While an initial DPIA was created, it was not updated as legally required, resulting in the deployment of a system inconsistent with the corresponding DPIA. Moreover, the recommendations and findings of this DPIA were not integrated into the updated version, which is typically a prerequisite for using a system with substantial risks. In contrast, the Corona Melder App continuously created and updated its DPIA, in compliance with GDPR requirements.

If the legislation regarding DPIA creation and updating had been followed, these systems would likely have needed to be reported to the Dutch Data Protection Authority (AP), which would have prohibited their use until privacy and security risks were mitigated. In severe cases, the Dutch Central Agency for ICT (AcICT) may have been asked to intervene, provide advice, and take further actions to improve the systems' success. However, this thesis cannot definitively determine the outcome of such a hypothetical situation.

There is some evidence supporting the theory that the Dutch government deliberately chose not to create these DPIAs to expedite the systems' deployment, despite the necessary changes to minimize data subject risks. Given that the Dutch government creates thousands of DPIAs annually for most, if not all, projects under its jurisdiction, it is unlikely that they lacked the resources or awareness to create one. The GGD and the Ministry of VWS were aware of HPZone's issues before the pandemic and still proceeded with HPZone and HPZone Lite without creating a DPIA. A data protection officer at one of the Dutch ministries suggested that the government might sometimes choose not to create a DPIA if other factors were deemed more important. In this case, the government may have prioritized rapid response to the pandemic over DPIA compliance, fearing that admitting the health systems' privacy and security risks would discourage individuals from using healthcare services during a crisis. This would also explain the lack of response to internal reports related to the lack of security within the GGD, as well as match observations shared by management-level individuals at one of the GGDs. However, this theory remains unconfirmed, and only a governmental inquiry could definitively verify it.

These kinds of situations are made possible by a general lack of transparency in project development and their corresponding organizations. The remarkable openness and transparency demonstrated by the Corona Melder App is the exception rather than the rule, with this approach not being applied to other projects. For instance, while any DPIA corresponding to an in-use system should not be able to contain any unmitigated risks, these documents are typically kept private and when released make use of excessive redaction of information. Neither is it easily accessible if DPIAs have even been created for any system, requiring WOO requests for any organizational information. This thesis has relied on WOO requests from Stichting ICAM and my own additional requests to access this information, as these requests are limited to the exact information that has been asked, to confirm that DPIAs have indeed not been created or created incorrectly. In the context of CoronIT, this was only possible because this system was ordered by the Ministry of VWS, as GGD GHOR does not fall under the mandate of the WOO. Without this, CoronIT would have been inaccessible to any investigation except for a governmental inquiry. In the context of HPZone, information was only accessible to WOO requests at each individual GGD, which could lack information that would only be present at GGD GHOR.

The same applies to the quality of CoronIT and HPZone's code and the flaws in their software. Since the project was not open-source, unlike the Corona Melder App, and in contrast to the government's intentions since 2020, the public had no way of learning about these issues. A DPIA could have identified areas of risk, but as previously mentioned, these were either not created or improperly executed, and even if they had been, public access would have been unlikely.

While CoronIT and HPZone may have been employed during a crisis, this thesis finds it improbable that the crisis situation itself was the primary cause of the numerous issues encountered. Time pressure may have played a role, but it should not have hindered the implementation of minor adjustments that could significantly improve these systems. For instance, no level of time pressure justifies the failure to activate monitoring functions or the lack of data storage for more than a month. Similarly, time constraints cannot explain the absence of protocols to review logged data or the inability to disable an export functionality accessible to all employees for an entire year. If the public had been informed about these issues before the major personal data breach in January 2021, the problems and resulting consequences might have been less severe.

9.2 Common and Open-Source Framework Vs. Crisis Specific Framework

Given the fact that crisis situations are unique and difficult to predict, it is difficult to adequately prepare and develop systems in advance. Therefore, the only realistic way to adequately prepare for crisis situations is to ensure that any given project is able to be quickly developed, to address any kind of crisis scenario. This thesis has chosen to address this through the development of a framework that offers a high level of flexibility, allowing for a variety of projects to be built, and offers the ability to develop projects quickly. In addition to this, it is important that any such framework should not reduce the level of privacy and security of any system and no shortcuts should be taken to enable the quick development of any such system.

Based on these facts, this thesis has chosen to create a common and open-source framework that is used for general situations as well as crisis situations, instead of a crisis situation-specific framework. The primary reason for this is that, in the creation of a framework flexible enough to be used in any crisis situation, the same framework would be flexible enough to be used to develop regular projects as well. In addition to sharing this overlap, the development of a common framework, and the significantly increased level of use, would make it significantly more likely for any such framework to be successfully adopted as will be discussed in further detail in this section.

The first major obstacle to the development of the proposed FAIR-based framework is related to economics. While the development of such a framework will require significant investments, investing resources to prepare for future situations isn't popular and it will be difficult to maintain the level of resources required to make this successful. In contrast, investing resources in a common framework that sees constant use is significantly easier to justify. A framework that sees constant use would see significantly higher levels of resources allocated to it. This would allow further development of the framework, making it more effective when it is used, either in a general or a crisis situation.

The second major obstacle to the development of the proposed FAIR-based framework is related to expertise. Closely related to economics, resources will need to be used to hire and pay for experienced personnel that would develop the framework. Considering the fact that the Dutch government already has a shortage of technical employees, reducing this number even further to create a framework in preparation for crisis situations that may happen in the future is an unlikely scenario. In contrast, using technical employees to develop a common framework to improve the general state of IT in the Netherlands is a far more likely scenario and matches the intentions of the Dutch government. By making projects easier to develop, standardizing data usage across entire sectors, and making personal data more private and secure and less likely to result in personal data breaches and political incidents this goal can be achieved.

This same obstacle also applies to the Dutch government's reliance on outside contractors. Given that the proposed FAIR-based framework is an entirely new standard, with contractors having little expertise related to FAIR given its age and less prevalent use outside the academic sector, expertise cannot easily be found. While the Dutch government is big enough to offer up a profitable market for any kind of standard they specify, a more limited and crisis-specific framework, would reduce the future number of contractors with expertise compared to the adoption of a more common framework.

The third major obstacle to the development of the proposed FAIR-based framework is related to capability. Considering that crisis situations aren't common, and each situation is unique, a framework that would solely be used in crisis situations would see little use. While the development of such a framework is already unlikely at best, based on the previous obstacles, such a low level of use would prevent individuals and organizations from gaining expertise in developing projects based on the framework. Without this expertise, no matter how well-constructed the framework is, any project developed based on the framework has a drastically lower probability of success. While continuous test projects could be created to try to establish this level of expertise, this faces the same obstacles related to economics and expertise, making it unlikely that tests will be conducted in the quantity required for this approach to be effective.

The fourth major obstacle to the development of the proposed FAIR-based framework is related to efficiency. In addition to gaining of expertise, a common framework that sees constant use would also result in numerous successful implementations of systems and functionalities, across a variety of projects. In a crisis situation, these previously created projects have significant value for the development of new projects, being able to be used as either the starting point of a new development or to reuse specific implementations of functionalities. However, this concept is only effective when there is a variety of previous projects available, which is unlikely for a framework that is only used for a limited number of crisis situations.

The fifth and final major obstacle to the development of the proposed FAIR-based framework is related to supervision. The level of oversight in projects is a major issue found in the GGD case study, with organizations such as the AP and the AcICT not being, or not being able to be, involved with the supervision of projects. While the AcICT may not have the resources or intention to evaluate every project they are mandated to evaluate, a common standard would address this by reducing the time it takes to evaluate projects. With a common standard, the AcICT would only have to evaluate the common standard itself, and any project-specific changes, which are likely to be minor and thus require less time. This also addresses the main goal of the AcICT, which is to improve the general state of governmental ICT in the Netherlands, as the common standard would apply to a larger number of projects. A framework that is only used in crisis situations would not have this same level of importance, making it unlikely that it would be prioritized over the evaluation of other projects.

9.3 Proposed FAIR-based Framework

The proposed FAIR-based Framework consists of three main areas of improvement, of which the first and the last areas will be discussed in detail in this section. The first area is the Dutch governmental ontology, which increases the level of transparency from the organizational layer down to the specification of the variables used in a project. Creating the ontology itself was beyond the scope of this thesis, instead opting for a description of the content and design of the ontology, to aid in the development of the Dutch governmental ontology, by the Dutch government, in the future. The third and final area is the common system framework, which allows for the easy deployment of either a default version or any other previously created projects based on the proposed FAIR-based framework.

9.3.1 Dutch Governmental Ontology

The creation of the Dutch governmental ontology used to support the proposed FAIR-based framework is the easiest to develop from a technical point of view, requiring a comparably small amount of development time and can easily be expanded through the addition of additional relations to other entities and information. Although it was beyond the scope of this thesis.

Collecting and storing this information will be a far greater change however, as while much of this information has already been recorded, this has been recorded over a large number of different sources and a variety of different formats. However, this primarily applies to large projects with an ICT component of more than 5 million euros, that have been developed or are in development around 2017 and later. Smaller projects, and older ones, especially ones that are considered legacy systems, have also not completed a DPIA, resulting in this information not being readily available. Large organizations can also have a large number of systems that would need to be recorded, such as the “Belastingdienst” which utilizes more than 800 systems, which would be a significant task to successfully investigate and store.

Storing this information and providing the public with access can have significant value as it increases transparency in government, but the information exposed to the public should likely be limited. In the creation of this ontology, it will create a comprehensive overview of all projects, but given the general state of ICT in the Netherlands, many of these projects may not be up to modern security standards. In this event, the creation of a comprehensive overview would be damaging to privacy and security as this would create an overview identifying valuable targets for malicious actors, which can then be used to determine the ones with the weakest level of security.

Based on this information, whether made public or not, significant investments should be made to address these flaws and ensure that systems are protected. This is both a requirement of the GDPR and as this information is stored somewhere, it is possible that a malicious actor would gain access to this list, which would expose the information even if it was not made available to the public. Which would increase the need to address these flaws even further.

The overall value for this specific component, even if the other components of the framework are not implemented, is significant. Creating a comprehensive overview of actionable information that currently simply doesn't exist. Allowing for automatic analysis of aspects that would previously have taken a significant number of manhours, or be impossible to do in general.

For example, the information included in the ontology would make it possible to automatically assess the general state of ICT across the entire Dutch government, based on the implementation of security measures, the creation of a DPIA, or even the number of projects built using the proposed FAIR-based framework. This information can then also be compared between various ministries, categories of data, or even specific organizations to identify and address any potential issues.

Another example is the ability to review project developers based on previous performance based on a wide variety of factors, to assess which vendors would be suitable to develop any specific project. For example, in the context of the case study, such an analysis would have identified that GGD GHOR is not the traditional developer of healthcare projects, and has a small budget, with the tender being a significant share of that budget, making it exceedingly less likely to be chosen compared to a different organization. If GGD GHOR was the only organization offering to undertake the project, the same approach could have resulted in a notification that this choice would have a lesser chance of success, potentially being connected to the supervisory organizations as the threat to privacy and security is high and the probability of success of the project without involvement is low.

Another example is the ability to more effectively prioritize which projects should be further investigated by supervisory organizations. The current approach of the AcICT, a sampling-based method is unlikely to be the most effective approach. If the purpose of the AcICT is to improve the general state of ICT in the Netherlands by sharing findings so that all projects can benefit, then projects should be chosen to maximize this effect. Either projects that are different in some regard from other ones to address unique issues or projects that are most likely to have problems that other projects would also have to have the most impact. Additionally, if the purpose to improve singular projects is the most important, then projects should be selected based on the size of their budget, to result in the most significant cost reduction. Or it should be focused on projects with a low probability of success to intervene in the projects that need intervention the most. Or it should be focused on projects that process the most sensitive data for a larger group of data subjects.

The current approach of the AP is to take action either when they receive a DPIA or when they receive actionable reports. However, this approach isn't effective when an organization decides to not create and/or report a DPIA or when no such reports are made specifically to the organization while problems still exist. To address these problems and more effectively utilize the resources of the AP, or even to request additional resources, the information contained in the ontology could be used to identify projects that would most likely benefit from the involvement of the AP. For example, legacy systems are not required to make a DPIA while they may process sensitive data without security measures. Currently, there is unlikely to be an overview of this system, which is why creating such an overview is so important. With this information, the AP would be able to warn each of these organizations that they need to make adjustments to these systems, and even though this recommendation is voluntary, it is ill-advised to ignore this advice as any incident based on the refusal of this advice would reflect badly on both the organization and the people involved with these decisions. Any other system that processes sensitive information, without the creation of a DPIA, or with the creation of a DPIA but without the implementation of certain security measures, could also trigger an investigation by the AP.

9.3.2 Development and Deployment

The proposed FAIR-based framework developed in this thesis is tailored to the specifications and requirements of the Dutch healthcare system, but it is not inherently unique to the Netherlands. Since the GDPR applies to the entire European Union, a system compliant with GDPR in the Netherlands can also be deployed in other member states, with minor adjustments for specific local laws. Moreover, countries outside the EU typically have lower data protection standards, and many are aligning their data processing laws with the stringent GDPR. Consequently, a GDPR-compliant system is likely to be compatible with local data processing laws in non-EU countries, although minor changes may be needed.

While the FAIR-based framework may be legally suitable, its practical implementation across countries is also crucial. The Netherlands is a wealthy country with considerable technical expertise. However, the framework does not require cutting-edge technology or extensive resources, making it feasible for countries with less wealth or expertise to implement it in their healthcare systems. The success of the VODAN Africa project, a FAIR network built with minimal resources, in Africa, supports the claim such a FAIR-based framework could be constructed all over the world. It should however be noted that VODAN Africa was able to benefit from a substantial number of partners, which is unlikely to be the case for other projects.

This is addressed however through the development of a common base that can be used to further develop and deploy projects. As this would be developed ahead of time, this would reduce or eliminate the significant technical requirements of developing a project based on FAIR. Additional parties could also be contacted and used to ensure that this development would result in a usable common base. Other countries could then make use of the developed common base, as well as continue its development in their own country. Through this, the Netherlands could then also benefit from the created improvements and projects by those different countries and organizations.

While the FAIR-based framework would offer a privacy and security-oriented way to develop projects, this does not necessarily address the unique nature of a crisis situation. Ideally, any crisis situation would already be addressed through a technical project advance, which is unlikely given the fact that any crisis situation is unique and likely to require its own unique approach. As such, any project to address any specific crisis may likely only be constructed during the occurrence of a crisis, at which point the system would already be required. This thesis has proposed the creation of a common standard that could be used to quickly construct such systems based on the principles discussed in the previous section, which would be addressed this limitation. However, the difficulty of developing such a common framework also exceeds the difficulty of building any individual project based on the FAIR-based framework. Making the creation of such a common framework difficult and requires a significant amount of support.

Another aspect related to this common framework that needs to be discussed is the fact that while there are many advantages to such an approach, such as an increased number of resources as resources can be re-used, a higher level of expertise with the common standard, and the standardization of the development of any project, this is not without risk. If a large number of projects were to be built on this standard, they would also share the standard’s security vulnerabilities, increasing the amount of data that could be breached when such a vulnerability would be discovered. While the probability of such vulnerabilities would be minimal given the number of resources a common standard could receive while remaining economical, this is only when the standard remains properly supported. If at any point the Dutch government, after the standard’s adoption, would decide to reduce its investments in the standard, the probability of such vulnerabilities would drastically increase, potentially affecting any project built based on the same standard.

Additionally, the FAIR-based framework aims to enhance transparency and public involvement in development. While this aligns with the Dutch government's intentions, the GGD case shows a lack of transparency, suggesting that some governmental organizations may oppose this trend. As the approach discussed in this thesis surpasses the intentions of the algorithm register and the intentions behind the development of the Corona Melder App, this may decrease the likelihood of the standard's adoption even further.

Increased transparency and open-source development can reduce mistakes and flaws in systems but also highlights them. If these flaws are not addressed quickly, they may lead to negative public perception and political repercussions. For example, if HPZone (Lite) had adopted the proposed FAIR-based framework's level of transparency or followed DPIA creation rules, security vulnerabilities, and risky design decisions would have been discovered more rapidly. This could have deterred people from interacting with the GGD or led to a political scandal, reducing the likelihood of the standard's adoption.

10. Conclusion

This study aimed to determine to what extent the FAIR guidelines can enhance GDPR compliance within governmental healthcare systems during crisis situations. To fully comprehend the findings of this investigation, it is essential to first understand the various components making up this investigation and research question. The first component, ‘FAIR guidelines’, relates to the flexible FAIR principles that can be implemented in diverse ways, as long as the core principles are adhered to. The second component, ‘GDPR compliance’, relates to the privacy and security of personal data processing in any kind of system as well as the additional data ownership-related rights data subjects have under the GDPR. The third component, ‘governmental healthcare systems’, relates to systems processing highly sensitive (medical) information, demanding rigorous security standards and protocols to ensure the privacy and security of this processing to align with the requirements of the GDPR. Given their critical role in public health, healthcare systems carry additional importance compared to other systems, as any potential problems could have implications beyond individual data subjects. For example, endangering public health as a whole due to either a lack of information availability or distrust in and avoidance of the healthcare process. The fourth component, ‘crisis situations’, relates to an unexpected and unique situation characterized by the inability to either anticipate or prepare for. Combining these components results in an investigation focused on applying the FAIR guidelines to address data privacy and security concerns within systems processing highly sensitive information during unpredictable crises.

In investigating this question, it has become evident that developing a solution exclusively for crisis situations would be either ill-advised or impractical. Such a solution would likely see limited usage compared to those designed for general situations, resulting in reduced resource allocation and expertise availability. Consequently, the overall probability of success would significantly decrease. Given the unforeseen and unique nature of crisis situations, any effective solution must prioritize flexibility and a high level of customizability, which would result in a solution that could be used for the development of many different projects. Therefore, there would be no reason not to make use of this solution in the development of projects beyond a crisis situation. Using this approach, any solution would also be more effective in a crisis situation, establishing a widely supported standard with substantial investment and expertise gained through constant use in various situations.

Any solution, regardless of its design, will inherently fall short of guaranteeing complete GDPR compliance for a system. While technical and organizational measures can be specified and made more accessible for implementation, achieving GDPR compliance remains project-specific, with the potential for mistakes to occur in any project. Even with increased transparency, the likelihood of detecting mistakes may rise, but there is no assurance that they will be promptly detected and addressed. Furthermore, certain GDPR violations may not result from mere mistakes but rather deliberate decisions, driven by a desire to enhance efficiency or expedite system releases. Transparency can contribute to revealing such decisions, but it cannot ensure their prevention or timely resolution. The most any solution can achieve is an improved level of GDPR compliance through the specification of technical and organizational measures to bolster security, while also facilitating the detection and resolution of both inadvertent errors and intentional decisions.

Furthermore, numerous smaller violations of the GDPR are often not related to technical or organizational measures. These violations can involve minor data breaches affecting only one data subject. For instance, a simple example would be the inadvertent delivery of information intended for a specific individual to a different recipient, such as when a letter or package is sent to the wrong address. The underlying causes of such incidents can vary, ranging from errors by the postal service to individuals providing incorrect information or data duplication errors. While potential solutions, like utilizing the basisregister Persoonsgegevens to eliminate data duplication errors, could be implemented, this investigation did not specifically address such minor violations.

The application of the FAIR guidelines in this thesis, following the proposed FAIR-based Framework outlined in Section 8, aims to enhance GDPR compliance in three key areas. Although the proposed Framework is purely theoretical, it does not rely on exotic or advanced technology, making it feasible to construct with resources beyond the scope of a Master's thesis. The first area of improvement involves creating a Dutch governmental ontology, for which thesis has described the content and design to aid the creation of the Dutch Governmental Ontology in the future. This future ontology expands the existing FAIR ontology for Dutch governmental organizations to include a projects and projects variable layer. The projects layer consolidates information related to projects that currently reside in numerous data sources, while the Projects Variable layer combines the traditional use of the FAIR principles, to standardize data formats and common terminology, with GDPR-related information. The main purpose of this area is to increase transparency and allow the use of information that currently would not have been easily accessible and/or queryable, to shape decision-making.

This area of improvement also has substantial value for the general state of Dutch ICT as a whole as due to the increased amount of information in a well-structured overview, it allows the ability to compare and monitor organizational and project-based information. Through the previous results of organizations in project development in any specific context, a better judgment can be made regarding which organizations should be in charge of developing any specific project. At a greater level, when specific information is added, it can function as a way to determine the implementation status of any governmental initiative in projects and organizations, allowing for both reports, comparisons between organizations, and the ability to more effectively address issues. The information contained in systems themselves, especially concerning their sensitivity and other GDPR-related information in combination with the implementation of security standards and GDPR-related procedures, can also be used by supervisory organizations to more effectively use their limited number of resources compared to their sampling-based approach. Allowing for either an earlier intervention, or increasing the probability of success in the most important projects.

The second area of improvement leverages the to be created governmental ontology by specifying the scope of information accessible to individual employees based on their roles, the specific data variables, and the context. This step aligns with the core GDPR principles of privacy and security by default and by design and data minimalization.

This area is especially important in the context of systems used by a large number of individuals, that lack expertise and a connection with the healthcare organization, as is likely in the event of a crisis situation. While systems used by a limited number of individuals, that are highly experienced, could be regarded as secure even when a significant amount of information is exposed to the user, this approach to data processing should not be the default for the development of systems. Any such system should therefore structure the access to data around the actual requirements of the user instead of the full data entry value. While the full data entry value could still be revealed to users, this should be done according to this structured approach instead of being the default. This way, the potential for data breaches from internal employees is significantly reduced or even eliminated.

The third area of improvement centers on establishing a common base framework that facilitates the development and deployment of FAIR-based projects. This concept is integrated with the Dutch governmental ontology, as the project-specific information is added to the ontology, and makes use of the second area of improvement to limit information exposure to reduce the risk to the privacy and security of the data subject. This is likely to be a critical requirement for the development of projects using FAIR, given the lack of experience with FAIR within the Dutch government and any possible contractors. Making it easier to develop and deploy projects would also significantly aid in crisis situations, where a quick but privacy and security-oriented development is required to develop a GDPR-compliant system.

Continued adoption of the common base framework improves the entire framework and every project built based on the framework, based on the ability to commit significantly more resources. With more projects being built based on the framework, other projects could benefit from re-using certain aspects or using these projects as a starting point for their own project. The increased level of experience with a single framework will also allow the Dutch government to be less reliant on outside contractors, while also increasing the probability of success even if outside contractors are hired as the government is now able to clearly communicate their requirements and monitor that these requirements have been implemented correctly.

However, it is important to note that the aforementioned improvement is dependent on correct implementation. If the findings from this case match the general situation, it is unlikely that all elements of the framework will be implemented correctly. While the introduction of even some of these elements may improve the current situation, it is important to recognize that it also introduces a major flaw. The use of any element of this framework results in a more secure data processing method, but an incorrect implementation may give actors the false impression that the entire system is secure because they have ‘applied the FAIR principles’.

One critical side note is that while the increasing adoption benefits the entire network, this can also cause a potential security vulnerability. With a high number of projects being built on a common standard, these projects will also become vulnerable to the same flaws. While this can easily be managed through the increased number of resources being allocated to the development of the framework, the danger resides in if these investments are ever decreased. If the government ever decides to stop supporting this framework, reducing the number of security updates addressing security vulnerabilities, the entire network will become vulnerable, risking a series of the biggest personal data breaches in Dutch history.

Ultimately, the FAIR principles and the proposed FAIR-based Framework presented in this thesis hold significant promise for enhancing GDPR compliance, but their successful implementation requires substantial commitment and support from the Dutch government. The first area of improvement is easy to implement on a technical level but requires significant organizational support to actually add the required information. The second area of improvement is easy to specify, which would at least ensure that recommendations can be provided, but offers no value if organizations are unwilling or not mandated to implement them. The third area of improvement requires a substantial number of resources to initially develop, after which the development costs are equal to or lower than current development costs due to development being shared over many different projects. As this approach aligns with governmental objectives, with there being significant benefits, the development of this area becomes more likely but it cannot be guaranteed.

11. Future work

This section explores the future direction of the proposed FAIR-based framework discussed in Section 8, along with its potential to encompass scientific data analysis. Section 11.1 outlines future work that may be done to aid in making use of the significantly increased amount of information contained in both the Dutch governmental ontology and the various systems they are connected to. Section 11.2 outlines the future work necessary to establish the feasibility and practicality of utilizing this framework within the governmental healthcare sector. Future work may also be done to explore the use of the same principles used in the FAIR-based framework to allow scientific data analysis without comprising the privacy and security of the data subject, which has been outlined in Section 11.3.

11.1 Usage of Automated Tooling

While the proposed FAIR-based framework would result in a significantly increased amount of information, in a structured format, making full use of this data would still require a certain level of technical expertise. Given that many individuals lack this experience, additional technological aid could be used to make the information contained in the Dutch governmental ontology even more accessible, and even allow data contained within systems to be more easily processed and used to fulfill various purposes. This section discusses future work that can be done in aiding the creation of queries, in Section 11.1.1, and utilizing large language models trained on this specific domain, in Section 11.1.2.

11.1.1 Aiding the Creation of Queries

Making full use of data, beyond having a well-structured overview, requires the creation of queries. Currently, most experience technical individuals will have will be in the context of current data standards using for example SQL. As ontologies based on RDF use a different format of queries, using for example SPARQL, this knowledge will first have to be gained by these technical individuals which would act as an initial barrier to make full use of data. Additionally, most people lack even these technical skills, which would prevent them from making use of this data at all, unless more easily accessible dashboards and interfaces are created to create these kinds of queries for them. Based on this, there is substantial value in being able to create queries, without users necessarily interacting with the query language itself. This section will discuss two approaches in which this can be addressed, which are the creation of these more accessible dashboards and the usage of an interpreter that can convert text into these kinds of queries. The second approach is discussed in further detail, with specific use cases, in Section 11.1.2.

The first approach is to create some kind of dashboard or interface that can be used to create queries while avoiding having to interact with the query language itself. This can be achieved by creating structural elements in which variables can be selected, in addition to specifying specific filters and transformations which can then be used to formulate the queries in the background. Using this approach, a user can write queries, without needing to have any experience with the query language. Based on the structured nature of ontologies, it should be possible to create this automatically, based on the data included in any data set. Either as part of querying the Dutch governmental ontology at the overview level or within the data sets themselves. This way, it would no longer be required to create dashboards and interfaces on specific (parts of) datasets, but a previously existing and elaborate mechanism could be used instead automatically. Of particular importance would be the ability to more easily create groups based on various conditions, and then use these groups to either combine these groups or use them as filters to find either joining or disjointed data.

The second approach uses natural language processing as a more easily accessible, although potentially more limiting, way to be able to create queries without having knowledge of the querying language. The limiting factor is the interpretation ability of the natural language processing model, which can be expanded over time to increase the complexity of sentences being able to be interpreted and the complexity of the queries being able to be generated. This is already possible through various large language models such as ChatGPT, Google Bard, and Bing AI chat, but by both integrating either one of these or a similar model into systems and by training the model on the Dutch governmental ontology and the information included in systems, this would be easily accessible to users. While this would require a more substantial investment, it is technologically feasible.

11.1.2 Usage of Large Language Models

Based on the recent significant advances in the area of large language models, it may now have become technically and economically feasible to apply this technology to make it easier to interact with the significantly increased amount of information that would be provided by the Dutch governmental ontology. Which would require the training of either new large language models are the expansion of current ones to create a large language model with domain knowledge of natural language, query creation as specified in Section 11.1.1, and information and descriptions from the Dutch governmental ontology. This approach has benefits far beyond the context of Dutch governmental health systems but could be applied to any sector and any form of information, as long as this information has been included in the Dutch governmental ontology. The feasibility of this has been demonstrated by the creation of GPT-3, GPT-3.5, and the most current model GPT-4 [232], which can be used far beyond the context of natural language interpretation, in use cases as diverse as the writing of code, debugging, and the creation of queries in both SQL and RDF. This section describes various ways in which such a model could be used.

In the context of the Regulatory use case by governmental officials, the application of these models could reduce the reliance on data engineers with significant expertise in the creation of queries. While the creation of such queries could be made simpler via the creation of a more easily interpretable interface, this approach would go one step further replacing this selection with natural language interpretation. For example, if a government official would be able to simply write a sentence describing exactly which information he requires, the model could translate this to an actual query and return the results of what normally would have been a complicated query. While this approach would be able to lead to the same result as one written by a data engineer with significant expertise, this approach is significantly faster, or would even be possible at all, versus the current situation where there is a clear lack of IT personnel in the Dutch government. An additional step would be to convert these results into various graphs, for example, the creation of a timeline where the implementation of governmental initiatives could be visualized over a certain period of time. Such a graph could then be used either by organizations themselves or could be used in support of political actors using this information to shape policy.

In the context of the Data Science use case by data scientists, the application of these models could make it easier to move from an initial investigation question to an investigation outcome. The investigation question would be used to first of all identify relevant information sources based on their description in the Dutch governmental ontology. After this, the model would interpret the information needed by the user and create queries that would send a request to the relevant parties and return the results of this investigation. When the investigation is simple, and information is directly accessible, this approach would allow such research to quickly complete or function as an initial investigation which can then be used for further research. When information is not directly accessible, this approach would also have significant value as it could communicate to the user that they would have to request access to these data sources. Based on the capabilities of the large language model, and the creation of a structured approach to request this access, it would be technologically feasible for this process to be done by the model itself. Which would then be able to further process the query and return the results at a time when such access has been provided. Similar to the previous use case, the results of these queries could be visualized across various graphs such as a country-level overview of the required information, a timeline, or a combination of both. With differences between periods being returned as well.

In the context of the informational need of citizens, who would most likely not have the technical expertise to write such complicated queries, or would lack experience with using structured information, the large language model would make it relatively easy to access any kind of information quickly. In many instances, information is available, but to make use of it would require the creation of interfaces and simplistic dashboards for there to be any way for the average citizen to interpret this information. Using this approach, such additional work is no longer required, and would instead be achieved through the development of a single model. Although this thesis is not able to estimate if this would reduce the number of resources required. While it would reduce the requirements to the creation of a single model, this model would be incredibly complicated by nature and require significant investment, while the creation of dashboards and interfaces would require few resources individually, but may exceed it through the sheer amount of data for which this would have to be created.

In the context of the need for information about any specific individual, this approach would make it significantly easier to interact with the Dutch government, for any possible need they may have. This will be illustrated through two different use case examples, with the first being used to get information from the Dutch governmental ontology itself, and the second being used to get information from specific systems about their situation. In both instances, there may be substantial value in adding the additional technical component of speech-to-text in the interaction between the user and the model for the request, and text-to-speech in the interaction between the model and the user for the results. This would make this useable for individuals who are visually impaired and individuals that are unable to interact with a computer. While this would still require the individual to be able to access the specific portal where voice interaction can be used, such interaction would be significantly reduced compared to situations where the individual would be required to type and visually interpret the complete process. With advanced in text-to-speech, making use of AI, natural-sounding speech can be created to make such an interaction resemble an interaction with an actual human.

In the context of the overall view of the Dutch governmental ontology, this approach would make it simpler to identify systems as this could be done via a general description of the system instead of more detailed knowledge. For example, a user could simply state that they want to view information about the system storing Covid-19 vaccination information, which could then be further narrowed down by additional questions instead of being presented with a list of all systems that process Covid-19 information or requiring the user to be aware that such a system is managed and used by the GGD of their specific region. Based on this approach and the information included in the Dutch governmental ontology related to specific projects, a user could then also use it to find out how their data is stored, which security measures were taken, and any other information that has been included.

In the context of getting information specific to the situation of the user, this approach can be extended to gathering the data within the systems themselves. Using DigID to authenticate the identity of the user, such information could be provided to the user securely, although this relies on the projects and systems that have been created to have implemented a standardized way to do so. The ultimate goal of this approach would be for users to have a single point of contact with the Dutch government to be able to accomplish any task, without needing to know which organization they would need to contact. Being able to use the functionalities provided by any such system. As an example of such a use case, a user may want to know which benefits they could apply for, however, they have neither the knowledge of which benefits exist nor which organizations they would have to contact to apply for these benefits. A sufficiently advanced large language model, and significant interoperability between systems, would make it possible to provide the user with exactly which benefits they could be eligible for to improve their personal situation and even be able to directly apply to any of these. Another, simpler, example would be for the system to be able to answer general questions such as what taxes there are on a specific income, or even be able to get the exact figures for the data subject and be able to easily explain how this is calculated.

11.2 FAIR-Based Framework

This section outlines the investigations required to determine the potential value of the FAIR-based Framework within the Dutch governmental healthcare sector. The first investigation entails exploring the feasibility of constructing such a platform, first in a general situation and then in a crisis situation, as discussed in Section 11.2.1. Subsequently, it is necessary to demonstrate the privacy and security by design aspect of the platform and compare it with other existing systems, as discussed in Section 11.2.2. Assuming that the platform can be constructed and provides enhanced benefits compared to conventional systems, it is imperative to investigate the framework's high level of customization to determine the most suitable implementations of each component in different circumstances, which is discussed in Section 11.2.3.

11.2.1 Demonstrating Feasibility

The proposed FAIR-based framework has been designed in a manner that eliminates the need for new or untested technologies, has general applicability to be used to support any crisis that involves data processing, and does not require significant resources. Despite these features, the framework remains a theoretical construct. The VODAN project serves as an indication of the framework's feasibility; however, it is imperative to note that the project is still in the development stage and has yet to demonstrate its ability to act as the primary support system for millions of people. Moreover, the development of the VODAN project was aided by several partner institutions, which may not be accessible to a platform established by or for the Dutch government.

Therefore, in order to demonstrate the feasibility of developing projects operating at the scale of the entire Netherlands based on the proposed FAIR-based framework, it is crucial to develop a prototype of that magnitude. The feasibility of the project could be initially demonstrated by constructing it without the constraints and limitations posed by a crisis situation, which would establish the framework’s overall feasibility.

In order to demonstrate the feasibility of projects based on the proposed FAIR-based framework, it is necessary to first construct the open-source and common framework that deploys various aspects of an initial system and includes some level of documentation such as a pre-created DPIA. However, the creation of such a mechanism poses a significant challenge, as the way the common framework is built heavily influences the projects built using it, and developing a mechanism that can be used for a variety of projects is likely to be significantly more difficult than the creation of singular systems that are based on the same FAIR-based framework. This poses a hindrance to demonstrating the viability of the framework in general.

Furthermore, the benefits of the common framework become more apparent with increasing use of the framework. Through this, both government employees and outside contractors would become more familiar with the underlying techniques and principles, which would increase the viability of the framework. However, without any previously created projects based on the framework, there would naturally be no similar projects, limiting the creation of any new project to using the first approach of project development as discussed in Section 8.6.1. Additionally, while this approach is the least standardized and most difficult way to develop a project using the proposed FAIR-based framework, this also cannot be mitigated by copying components from other systems, which would have been possible in a more mature and established implementation of the framework.

Upon creating the open-source and common framework, the next step would entail utilizing the framework to develop a project under the absence of constraints on time and resources. Although this scenario is not reflective of crisis situations, the primary objective of this initial test is to establish the effectiveness of the common framework and its ability to support the development of successful projects. It is important to note, however, that a failed test does not necessarily imply that the project is infeasible. Rather, it indicates that the specific project, with its level of expertise, was unsuccessful.

Depending on the outcome of this test, the next iteration of the project would learn from these errors and rectify them in the future, although it should be noted that the failure of any version may be viewed by some as a failure of the entire framework in general, which could result in the ceasing of all development on both the proposed FAIR-based framework. This presents a significant problem given the complex nature of developing a framework that can be adapted to different contexts, as well as the challenges inherent in project development. Particularly in light of the Dutch government's track record with ICT projects and their limited experience with the FAIR standard. Therefore, it is improbable that any first iteration would succeed.

The second phase of demonstrating feasibility involves showcasing the project's capabilities within the constraints and limitations of a crisis situation, after successfully building the platform. However, replicating a crisis scenario presents challenges as such situations are often unique, and many constraints and pressures cannot be duplicated in an experimental setting. For instance, the political pressure exerted on the GGD while creating the system cannot be replicated, yet it is highly probable that future systems developed during a crisis will face comparable political pressure. Furthermore, the platform's issues stemmed from the activities of a large group of employees who lacked adequate training or demonstrated malevolent behavior. Testing the platform with a similar number of employees would be prohibitively expensive and would not accurately replicate the behavior of individuals accessing information, whether by accident or for personal gain. Although activities can be simulated to test the system's ability to handle a particular level of activity, they cannot replicate actual human behavior.

Two constraints that can be replicated to imitate crisis situations are the limitations of time and budget. Tight deadlines can be set to simulate the time pressures that arise during a crisis. Similarly, budgetary restrictions can be simulated to test the project's capabilities in situations with limited resources. However, it is uncertain if all crisis situations are associated with budgetary constraints, as highlighted in one of the interviews with the GGD. During the Covid-19 epidemic, any budgetary restrictions the GGD faced were eliminated, suggesting that not all crises have limited resources.

11.2.2 Demonstrating Privacy and Security by Design

The proposed FAIR-based Framework has been developed in line with the principles of privacy and security by design, with the understanding that the security and privacy of the implemented system are dependent on its adequate implementation. While the use of the common framework can guarantee a baseline level of security, given that certain aspects of the system have been pre-created and proven to be secure, improper deployment or a lack of organizational security measures could potentially compromise the system's integrity. As a result, the presence of vulnerabilities could undermine the effectiveness of implemented protections, allowing for the circumvention of security measures. Therefore, it is crucial to ensure that the FAIR-based Framework is deployed following proper procedure and is supported by robust organizational security measures to maintain the privacy and security of the system's users.

In addition, regular updates are necessary for the maintenance of the common framework's security. Organizations must ensure that they apply these updates promptly, as known security vulnerabilities can be exploited by malicious actors if neglected. This is of particular importance to the validation engine, which governs the information available to users and therefore plays a critical role in the system's overall security. Any weakness in this component has the potential to undermine the system's security. Penetration testing, commonly used to evaluate a system's resilience to external attacks, remains a valuable means of verifying the implementation's security and assessing the adequacy of organizational measures to prevent security breaches.

Furthermore, it is essential to establish that the system's security protects against internal attacks. While the monitoring system aids in detection and ideally quickly identifies any attempt, preventing an attack entirely is always better. Based on the case study, internal attacks pose a significant threat, and prevention of such attacks is usually not tested during the development of new systems. The FAIR-based framework limits the amount of data each user can access and the amount of information displayed to the user, but only if proper procedures are followed.

By conducting a large number of attacks, both from the interface that employees have access to and any custom client that could be developed to circumvent security measures, the security of the system in the context of internal attacks can be established. This can be achieved by using a combination of known vulnerabilities gathered by organizations such as the Dutch Nationaal Cyber Security Centrum and the services of white hat hackers, who can be contracted through either a contract, a hackathon, or a bug bounty program.

The results of this investigation can be used to identify and address security vulnerabilities within the common framework or the specific implementation of a project, thus enhancing the security of current and future projects utilizing the proposed FAIR-based framework. Additionally, the results can be compared to those obtained from a security assessment of a traditional system, providing further evidence in support of the transition to the FAIR-based framework approach advocated in this thesis. However, such a comparison should be delayed until the proposed FAIR-based framework is more established and developed, as it could otherwise be misused as evidence against the framework.

11.2.3 Determining the Optimal Component Implementation

The proposed FAIR-based Framework has been designed with a specific focus on the fundamental components of a system, resulting in a structured foundation that can achieve the aim of enhancing data control while also reducing the likelihood and scope of a personal data breach. The framework offers a high degree of customization, which allows projects developed using the framework to be able to adapt to various data types and situations. In this thesis, the focus has been on describing how such a system could work, with future work being able to investigate a more concrete implementation.

In addition to this, while the common framework element of the proposed FAIR-based framework is valuable, a universal implementation may not necessarily be optimal for all contexts. Further work can be conducted to investigate a variety of systems across the healthcare sectors and other sectors, as well as possible crisis situations with varying data needs, to enhance the proposed FAIR-based framework and establish best practices for each situation. Additionally, a usability investigation may be conducted to determine the impact of increased security measures on system efficiency and output.

Section 11.2.3.1 outlines the future work that could be done to investigate to what degree roles and functions could be divided, taking into account the aspects of usability and security. Section 11.2.3.2 outlines future work that could be done to determine the optimal implementation of the monitoring system, especially regarding both the Trained Behaviour Engine and the list of individuals of interest. Section 11.2.3.3 outlines future work that could be done to determine the optimal implementation of the search system, especially regarding the specific threshold and the way that similar results are handled. Section 11.2.3.4 outlines future work that could be done to determine the optimal implementation of the detailed patient dossier system, especially regarding the specific threshold and conditions that determine how much information is exposed to the end user. Section 11.2.3.5 outlines future work that could be done to determine the optimal implementation of the aggregational statistics system. In the current version of the proposed FAIR-based framework, this system is quite limited however, with significant additions allowing for scientific data analysis without comprising the privacy and security of the data subject being discussed in Section 11.3.

11.2.3.1 Roles and Functions

The proposed FAIR-based Framework's methodology towards user roles and access to functionalities serves to enhance privacy and security by improving upon the traditional systems' permission matrix-based approach. In contrast to simply designating which functionalities employees have access to, this approach divides data and functionalities into smaller sub-roles to regulate access. Though employees would still have access to the same number of functionalities necessary for their responsibilities, each role would necessitate a separate login, limiting the number of functionalities that can be accessed simultaneously.

While this approach is feasible from a technical standpoint, further research is necessary to determine its impact on system usability. The subdivision of functionalities is bound to decrease employee efficiency since they must undergo a login process to access other functionalities. Moreover, selecting the incorrect role would result in further inefficiencies. The findings of this study could provide insight into identifying functionalities that could benefit from being subdivided into sub-roles and those that could remain combined. However, it is crucial to emphasize that although efficiency is important during crisis situations, security should never be compromised, particularly when handling highly sensitive information such as medical data.

Another potential avenue for research is this framework’s intention to have the user select a region to which they have access. While this approach provides significant security benefits, especially in combination with a properly implemented monitoring system that detects region switches, it may result in a loss of efficiency. To mitigate this potential inefficiency, an alternative solution could involve permitting users to switch to another region they have authorized access to without requiring a logout and login process. This solution would still be able to benefit from the fact that access is limited to smaller data sets in addition to the fact that any such switch is logged by the monitoring system but would still compromise some aspects of security. Consequently, further investigation is necessary to determine the most appropriate approach for specific scenarios. The framework's high level of customizability even allows for a user-level application of region-based access, enabling certain users and user roles to access larger data sets than others.

11.2.3.2 Monitoring System

The proposed FAIR-based Framework offers a novel approach to data access, providing a high level of control over data interactions and generating substantial information for monitoring systems to detect, deter, and prevent malicious activities. The framework's method of data management eliminates the need to transfer data externally, with few exceptions, thereby enabling organizations to monitor all data interactions effectively. This section discusses potential areas of future research that could enhance the security and privacy of the FAIR-based Framework through the monitoring system, such as a more extensive investigation of the Trained Behaviour Engine and the implementation of a list of individuals of interest across various sectors.

Future work could investigate the optimal utilization of the extensive monitoring data to ensure heightened security against internal malicious actors. Although Section 8.5.2 already outlines various applications of the data collected by the monitoring system, the Trained Behaviour Engine, as detailed in Section 8.5.2.5, merits a more in-depth examination. This component's primary objective is to differentiate between legitimate and malicious actors, going beyond basic detection methods such as elementary performance indicators, which primarily identify malicious actors attempting to bypass other monitoring system components.

An interesting approach, aligned with the one described in Section 11.2.2, involves engaging white hat hackers to attempt to circumvent the monitoring system, while developers iteratively strive to detect their behavior. However, this method presents practical challenges; while the core components of the common framework can be economically tested—benefiting all projects built upon this framework—project-specific changes would necessitate individual testing for each project. Although project-specific changes may apply at least partially to different projects. Despite the potential economic concerns, this approach represents a compelling area of study and could be incorporated into future projects to enhance project-specific modifications to the monitoring system.

Future research could also examine the implementation of a list of individuals of interest to flag search queries involving any person on the list. Although this system has proven successful in the healthcare sector, its adoption remains limited in other industries. Although it should be noted that even in the healthcare sector, not all systems have implemented such a component. Thus, investigating the development of a common standard for this system across various sectors could enhance the privacy and security of the listed individuals, a crucial consideration given the potentially severe consequences of personal data breaches for many on the list.

In the healthcare sector, the system's implementation is relatively straightforward due to the legal requirement of using the Citizen Service Number (BSN). Verifying whether a BSN is on the list of individuals of interest should not raise privacy or security concerns, provided that the implementation is carefully executed. A person's presence on the list may constitute sensitive information, although linking the BSN to a specific individual requires an additional breach. Nonetheless, given the prevalence of personal data breaches, this remains a distinct possibility. Therefore, any chosen solution should ensure that the list is well-protected and encrypted, with access restricted to authorized organizations.

The precise implementation of such a system also raises unexplored practical questions. Two primary approaches exist: each facility creating its own list or establishing a general system accessible to different organizations. While the former would eliminate the need for restricting access to authorized organizations, it would significantly diminish the system's utility, as individuals would need manual addition, contradicting the system's purpose. An automatic addition would necessitate a general list, mirroring the second approach. Consequently, future research should concentrate on the latter approach, involving a general list accessible to organizations.

Adopting a general list of individuals of interest presents significant advantages but also entails resolving other practical questions. For example, determining who is authorized to add individuals to the list and establishing the criteria for inclusion requires clarification. As an illustration, if government employees were to be added, would all qualify, and would they be removed upon leaving government employment? The issue becomes more complex when considering for instance celebrities, as defining who is or is not a celebrity proves challenging, particularly for social media influencers. For instance, if using Instagram, would a follower-based system be employed, and who would determine the necessary follower count for inclusion? Any potential implementation would confront similar practical problems.

Addressing the practical questions concerning list inclusion could result in an excessively extensive list of individuals of interest, diminishing the value of having such a list. If every query raises a flag, the entire system becomes ineffective. Different categories of individuals of interest could be created, but the initial problem of determining qualification criteria resurfaces. Although establishing a list of individuals is a straightforward concept and easy to implement, numerous practical questions must be resolved to ensure the list contributes to the system's efficiency and effectiveness.

Once the practical questions surrounding list inclusion are resolved, the next set of challenges pertains to list management. As the list is available to multiple organizations, a designated organization must assume responsibility for the list's security, technical upkeep, and access determination. As previously noted, not everyone can be granted access, as confirming a BSN's presence on the list involves sensitive information and should be restricted to authorized, secure organizations.

11.2.3.3 Search Requirements

The proposed FAIR-based Framework's approach to search queries, which involves setting additional search requirements and transforming the returned data based on the query's specificity to expose less sensitive information, can significantly enhance data subject privacy by eliminating a potential avenue for internal attacks. The amount of information exposed to the user can be limited using various methods, depending on the data's nature and the associated risk to the data subject. Furthermore, the returned information quantity can be based on the search query's number of results. This could be configured to return no information for overly generic queries, return limited information when the number of results exceeds a certain threshold, or provide more detailed information for increasingly specific queries. Employing a system that necessitates registrable user interaction for accessing more detailed information further bolsters the system's security.

Future research could examine the specific conditions and thresholds applicable to different systems and circumstances while weighing the usability and security of this approach. The most secure solution would only return results when the system identifies a single result; however, this method is inefficient and, consequently, unlikely to be usable. If conditions are less restrictive, the system risks resembling traditional search systems, which return numerous results containing personally-identifying and valuable information. A balance should be struck for general and specific situations, thereby establishing a set of best practices for various scenarios.

Future studies could also explore how this approach could be integrated with functionalities developed to enhance the search process, such as searching for similar results to correct misspellings or inaccurate information. This is particularly relevant to similar data entries, such as surnames – for example, "van der berg" and "van den berg" are close enough to be confused. The search functionality created for the proposed FAIR-based framework prioritizes security and does not account for this aspect. While implementing such a functionality would be technically straightforward, determining the appropriate implementation is considerably more challenging. Addressing the aforementioned example would either significantly increase the number of returned results or necessitate even more specific search queries through the use of additional conditions to decrease the number of results below the threshold again.

11.2.3.4 Patient Dossier

The proposed FAIR-based Framework's approach to patient dossiers, which are detailed records about a data subject, involves applying data transformations and functionality over data as described in Section 8.5.4, and increasing monitoring as explained in Section 8.5.2.4. This method significantly enhances the privacy of the data subject by removing another potential attack avenue and increasing the likelihood of detecting such attempts. In the FAIR-based framework, the exposed information depends on the sensitivity of the data, the user's selected role, and the requirements for their responsibilities. As discussed in Section 8.3.2.1, more detailed information can be revealed to the user, depending on their user role, following user interaction. It should be noted that neither the use of patient dossiers nor user interaction to access more detailed information is unique to this architecture, as some form of these techniques is common practice in parts of the healthcare sector. However, these techniques are not employed universally within the healthcare sector, as demonstrated by this case study, and their application has been expanded beyond the current implementation in traditional systems.

Future research could explore the specific conditions and thresholds that could be employed while considering both the usability and security of this approach. For data entries such as phone numbers or email addresses, these conditions are likely to involve displaying a function instead of any value, which can then be used to perform the function for which the data would be utilized. However, for other data entries, the transformations are less evident, and a balance must be found between usability, which would likely necessitate less restrictive data transformations, and privacy and security, which would likely require more restrictive data transformations. Although it is improbable that a general solution applicable to all systems can be discovered, the results of such an investigation could reveal a general solution for specific data variables or a series of steps that can be taken to apply the appropriate conditions and thresholds to individual projects.

11.2.3.5 Queries and visualization

The proposed FAIR-based Framework's approach to queries and visualization is the most similar to the approach used in more traditional systems while still containing significant differences in specific aspects. As long as organizations can ensure that data itself is impossible to trace back to any individual, providing access to this data does not pose any challenges to the privacy and security of the data subject. Organizations also already make use of transmitting results over raw data itself, for example in the context of a dashboard where the only matter of importance is to present accurate and recent numbers instead of allowing for any type of investigation.

The proposed FAIR-based framework can easily accomplish the same level of functionality via data visitation, however, the value of the proposed framework may be significantly higher. Instead of being limited to datasets uploaded by the government to sites such as 'data.overheid.nl', the data visiting approach, combined with citizens being granted a very low level of access to aggregational statistics, would allow every single system using the proposed FAIR-based framework to be a possible data source. Without any action from any organizations, except for configuring the level of access that would not result in a risk to the privacy and security of the data subject, any individual would have access to up-to-date and accurate information from a variety of systems.

Future work could investigate how such a service could best be made a reality, including what level of access should be offered, how extensive this service would be, and what the expected level of use would be if such a system were to be created.

While the technical aspects have already been addressed by the proposed FAIR-based framework, with the addition of users being as simple as creating a new user role and setting its level of access, a safe level of access for many different situations and different types of data could be determined in advance.

Relating to the extensiveness of such a service, there are many different possible implementations of such a system. Instead of simply offering up data sets with values, the proposed FAIR-based framework could be used for substantially more. Using the same advanced system that regular employees would have access to, while operating under vastly stricter conditions, citizens would be able to look through data without requiring any of the technical expertise that would normally be required for such simplistic investigations. If this could be combined with the possibility of making visuals similar to a governmental dashboard, identifying trends (time, location, various groups, etc.), this would increase the value of the approach even more.

Identifying the level of use is also important as this is a consideration for the standards to which any system is built. Offering up access to citizens would require more resources to be available, with a low level of use resulting in none to a marginal increase in cost while thousands of users would be serious consideration for how to build systems in the future to account for this fact. The level of use is directly related to the usefulness of the system and is difficult to anticipate for such a novel approach, which makes research into this aspect difficult but valuable.

In the future, the potential value of this approach could be increased even further by developing a way to make the proposed FAIR-based framework support Scientific Data Analysis as well, as discussed in Section 11.3.

11.3 Expansion to Cover Scientific Data Analysis

The proposed FAIR-based framework, applying the same techniques discussed in Section 0, holds significant potential for scientific analysis. However, we chose not to integrate this concept in the current version due to feasibility concerns. Creating a system that performs both functionalities without compromising privacy and security for the use case presented in this thesis would require additional research. Using this framework to support scientific data analysis would also require significantly more resources than the use case discussed in this thesis, which is likely to exceed those available to local health facilities, especially during a crisis situation. In addition to this, supporting scientific data analysis with this level of sensitive data poses societal and legal implications, even if the system itself can guarantee security.

Section 11.3.1 outlines the compatibility between the two use cases, considering the distinct requirements of scientific data analysis compared to the thesis use case. Although these challenges are not insurmountable, it is improbable that a single system can address them while maintaining the same level of privacy and security. However, this does not indicate that the technique itself would bring privacy and security in jeopardy, or that this approach would not be significantly more able to guarantee the privacy and security of data subjects over a traditional approach.

Section 11.3.2 outlines the unique complications that arise when applying the proposed FAIR-based framework to scientific data analysis. These complications are not related to the system's security or anonymity but are factors that must be addressed to maximize its value. Specifically, they pertain to the conditions under which society and the law would permit data access for scientific analysis. Although the techniques used can ensure security and the GDPR allows for exceptions in scientific research, addressing these complications remains essential.

Section 11.3.3 outlines two of the major changes that the data visitation approach would bring to Scientific Data Analysis. These changes are related to a change in the location and the number of resources that are required for Scientific Data Analysis as well as a change to the amount of data that would become accessible for scientific data analysis and the number of organizations that would benefit from this system.

11.3.1 Compatibility with the FAIR-based Framework for support of the healthcare process

The proposed FAIR-based Framework's compatibility with scientific data analysis poses challenges in at least three different technical areas, which will be covered in this Section. Section 11.3.1.1 outlines the additional functionalities that will be required to support the scientific data analysis process. Section 11.3.2.2 outlines the compatibility issues related to the increased number of resources that would be required to support scientific data analysis. Section 11.3.1.3 outlines the compatibility issues related to the source of the data, with scientific data analysis likely requiring data from multiple sources to be combined, which contradicts the principle that data can be stored at a single location and accessed via the principle of data visitation.

11.3.1.1 Additional Requirements for Scientific Data Analysis

The proposed FAIR-based Framework has been created primarily for data processing for supporting the healthcare process. Supporting data processing in the context of supporting scientific data analysis will require multiple additions, most of which contradict the principles of only being exposed to data when required and for organizations to be in full control over their data.

The first possible obstacle can be found in the pre-processing stage of data analysis. Instead of making use of the pre-existing infrastructure surrounding the use of existing variables, scientific data analysis will require the creation of new variables, derived from existing ones, either to make adjustments to existing data, to make new categories, or transform these into a different form entirely. While this is simplistic when organizations can be sent a data file for them to expand on this in their environment, this does not apply to the data visitation approach used in the FAIR-based framework. Instead, this would require the data source to create some kind of environment where other organizations can achieve this same level of functionality, while also ensuring that the privacy and security of personal data is never compromised. Future work could explore this, especially in combination with some form of virtualized environment, which could become a general expansion of the framework which can then be utilized for all other projects.

The second obstacle can be found in the code execution stage of data analysis. Instead of relying on the more simplistic models and algorithms that have already been incorporated in the proposed FAIR-based framework, this would require either a significant amount of development to incorporate all the possible models used in data analysis or the ability for organizations to use custom code to apply to the data. For the second approach, this will need to be combined with a way to prevent the presence of custom code from being able to be exploited to gain access to more data than the system would normally allow for. Future work could explore which of these two approaches is the most viable approach, although both with likely make use of some form of a virtualized environment. Although a manual approach where code would be sent to the organizations for review and processing may also be a possible solution. Successful research into this could lead to a general expansion of the framework which can then be utilized for all other projects.

11.3.1.2 Increased Number of Resources Required

Supporting scientific data analysis, either via the same hardware used to support the healthcare process or with hardware that has been allocated to scientific data analysis, will in most instances require additional resources. Depending on the complexity of the models required for analysis, as well as the size of the data set, this increase in resource demand can be significant and, in some instances, even exceed the resources required for the original system.

Future work could investigate how big the impact of the inclusion of scientific data analysis would be, at various levels of implementation. Additionally, it could be investigated if there would be any way to schedule research to take advantage of downtime beyond working hours, and how big of a positive effect this would have on the required number of resources.

Additionally, it should also be considered that by making such a system available, it may increase the amount of scientific research that is being done beyond the current level. Which would increase the number of resources required to support this system even further. Which is discussed in more detail in Section 11.3.3

11.3.1.3 Combining Data from Multiple Sources

The proposed FAIR-Based Framework presented in this thesis has primarily focused on scenarios where databases are treated as separate entities, with data visitation being used to allow users to access query results without any actual transference of data. However, this approach is not applicable in all scientific data analysis investigations, as some investigations require combining data from multiple sources. For instance, studying the excess mortality rate due to the Covid-19 pandemic, as discussed in Section 2.1.3, necessitates the integration of data from multiple sources.

This integration of data from various sources poses significant risks to both the organizations involved and the individuals whose data is being processed. First of all, the presence of data in multiple locations goes against the principle that data is only stored at one location. This complicates control over the data, as access can now be provided at two locations that each can have different conditions. The monitoring system is now required to register activity at multiple locations, which complicates acting on user activity and goes against the simplicity of all data access being centralized in a single system. The presence of data in multiple locations can also introduce errors when changes aren’t synced between locations.

In addition to these factors, there is also a significantly increased risk due to the heightened value of these organizations as targets for potential data breaches and the inherent risks associated with the combining of data from multiple sources. With more data being available, the potential value of any attempt to breach security is increased. The transfer process, as well as the need to share the decryption scheme with the other organization, increases the vulnerability of the system. As such, it is crucial to implement robust security measures to protect both the organization's and the individuals' sensitive information during such data integration processes.

To illustrate the difference between combining data from multiple sources in the proposed FAIR-based framework as discussed in Section 0 and in support of scientific data analysis, Figure 58 and Figure 57 have been created.

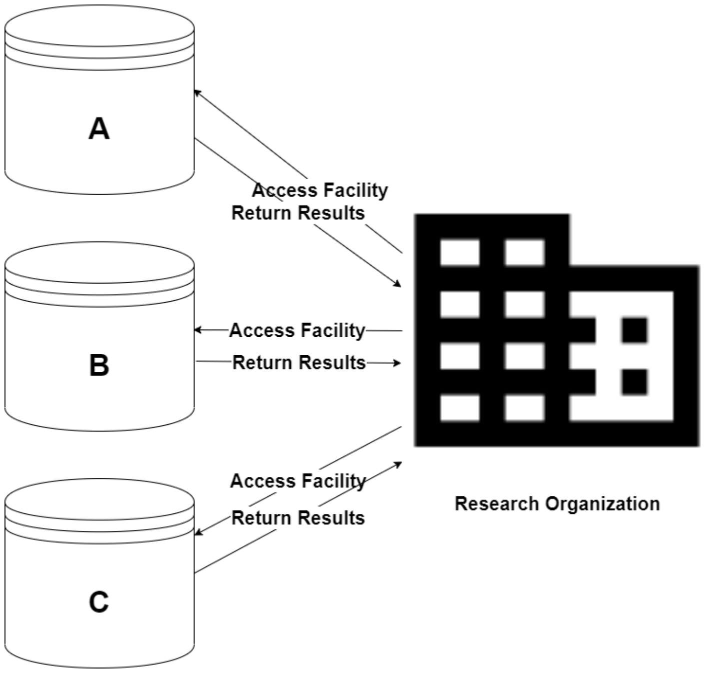


Figure 58 - Combining results from multiple facilities using data visitation

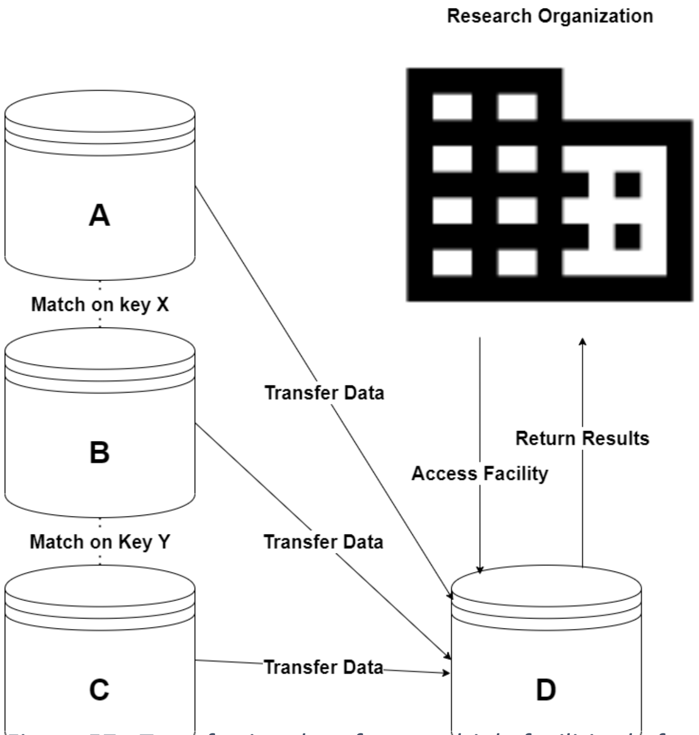


Figure 57 - Transferring data from multiple facilities before using data visitation

In Figure 58, the research facility can send a similar query to each data source, which then processes this request and sends the results back to the research facility. In this scenario, there is no transfer of data, and depending on the configuration of the aggregational statistics system, no personal data of any kind is exposed. This is an example of the application of the concept of data visiting.

In Figure 57, it is not possible for a research facility to simply send a similar query to each data source, as data from multiple organizations or systems need to be combined by matching a key. At some point during this process, a location needs to have access to the raw data from datasets A and B to match on a certain key X, and a location needs to have access to the raw data from datasets B and C to match on a certain key Y. In this example, this is done by transferring the data to dataset D instead of transferring data to datasets A, B, or C however both situations are technically feasible. While the initial transfer of data violates the concept of data visiting, data visiting could still be used to provide the research facility with its results from this dataset D.

Having examined both possible solutions, which are to either conduct the data combination at one of the organizations involved, reducing the amount of data that needs to be transferred and avoiding duplication, or to use a third-party organization specifically designed for data security, this thesis prefers the second approach. Based on the high level of security that is required to prevent any personal data breach, especially considering the increased value of combining information.

Future work could determine the feasibility of this approach and investigate what the security requirements would be for such a third-party organization. When the requirements have been created, an investigation can be done into viable options, or in the absence of those, outlining a new organization that can be created. Specializing in providing scientific data analysis for projects built on the proposed FAIR-based framework.

11.3.2 Addressing societal and legal implications

If the proposed FAIR-Based Framework discussed in this thesis were to cover Scientific Data Analysis, the main problem is not a technical one but an organizational one. Even if data would need to be transferred, thereby violating the goal of keeping data at the location that collects it and complicating control over the data, the transfer itself should not compromise privacy and security if done between secure organizations. At no point would information be shared with research organizations as they would be provided access, through data visitation, at a different location. Instead, the main problem with this approach is to use the data itself. It isn't enough to ensure that there is no risk to privacy or security, it's the perception of security and the legality of accessing the data that determine if such a system may be used.

The proposed FAIR-based Framework already offers up a system for data subjects to offer up their consent to use their data for a variety of purposes. This is achieved through the data subject portal that would either provide the data subject with an overview of all their data usage through the central registry as outlined in Section 8.2.4 or through a project-specific portal as outlined in Section 8.5.6. However, while this introduces a technical solution to offering consent, future work will need to be done to determine if and how a solution can be created that meets the requirements for informed consent as set out by both the GDPR and the "Wet medisch-wetenschappelijk onderzoek met mensen" (hence 'WMO').

Section 11.3.2.1 outlines the effects that the legal requirement of informed consent will have on the use of data collected and stored in any system built on the proposed FAIR-based Framework for use in scientific purposes.

Section 11.3.2.2 outlines the necessity to not only ensure that the privacy and security of the data subject are guaranteed, but to also convince the data subject itself of the security of this approach. Without this, it becomes unlikely that any data subject would be willing to give consent to use their data for scientific research, making any kind of implementation, no matter how comprehensive, ineffective at allowing for scientific research.

Section 11.3.2.3 outlines the fact that even in the absence of personally identifiable information, some variables may still contain information that can be led back to a singular individual or a small group of individuals. Future work will need to be done to minimize the probability of this and to be able to account for this in the proposed FAIR-based Framework.

11.3.2.1 Informed Consent

The concept of informed consent is important in the context of conducting medical studies, but this thesis was unable to investigate the exact relation between informed consent and the processing of personal data using aggregational statistics. When anonymized data is being processed, as long as there is no way to identify any specific individual, this does not fall under the GDPR and does not require informed consent. The aggregational statistics approach used in the proposed FAIR-based framework is also anonymous, with no personal information being exposed, but could potentially introduce legal questions given that this processing is done on the original data set.

The most likely instance where informed consent would be used in combination with the proposed FAIR-based framework would be in its deployment to support small healthcare projects conducting a medical study. While there is no reason why this framework would not be able to operate at this scale, it has not been the focus of this thesis, being focused instead on larger systems with a larger number of employees, where the focus on privacy and security becomes more valuable. Informed consent would most likely be conducted orally instead of digitally, reducing the value of its inclusion in the framework, although there would be some value in being able to electronically revoke consent using a simplistic interface.

If informed consent is ever required digitally on larger systems, it needs to first be investigated which types of consent would meet the legal standard for informed consent. The most optimal approach for the support of scientific research would be some form of general consent which would allow organizations to use the data for a variety of purposes without requiring additional consent. However, while this would aid scientific research, it would need to be legal to do so. Any other approach involving a more limited form of consent would be less valuable, primarily due to individuals being unwilling to continuously provide consent for each individual scientific research project.

Future work can investigate the exact circumstances where it is required to get informed consent beyond the exceptions that Article 89 provides. It can also investigate, in such instances where informed consent is required, in which form this can be provided, and how this system can be designed in such a way as to encourage providing informed consent.

11.3.2.2 Convincing Individuals of The Privacy and Security Aspects of This Approach To Data Management

As stated in the Literature Review Section, it is ultimately the perception of security that shapes human behavior over the actual security of systems. In instances where consent from the data subject is required to process their data, it is therefore prudent to investigate ways into how this perception can be influenced and how the positive effects on privacy and security can be made clear.

Additionally, it must be clear what exactly a data subject gives permission for, as it is not always required to provide permission depending on the nature of the data and the purpose of the processing. For example, article 89 of the GDPR provides exceptions based on the purpose of scientific research. If data is processed for this purpose without the data subject having given their consent, the processing itself may be entirely legal, but the resulting misinterpretation could lead to the data subject believing that the system itself doesn’t work properly. In this case, the data subject would be unlikely to provide consent for research that would not fall under this exception.

Given the fact that data is processed at the source of the information, with even these queries accessing the data at the source, this information may be of use in influencing the data subject. If a record of such queries can be shown to the data subject, including the exact purpose and ways in which the data is being used to help societal causes, the data subject would likely gain a more positive view of this process and be more likely to provide their consent to further contribute.

Future work could examine each of these factors to determine which strategies are effective in convincing people of the privacy and security benefits, and which strategies are effective in promoting the provision of consent to processing their personal data.

11.3.2.3 Seemingly Anonymous yet Distinct Information

It is important to consider that even seemingly anonymous information may yet contain distinct information that could potentially lead to the identification of the patient. The presence of a rare genetic disorder or a unique combination of medications and treatment can be enough information, even in the complete absence of a name, age, or BSN to potentially be traced back to either a singular individual or a small group of individuals. Additionally, cross-referencing information with other data sources, for example, the date of treatment or a location with other data sources would lead to the identity of the patient of be exposed, even in the absence of usual identifying information.

Future work needs to be done to establish which seemingly anonymous information should be treated with more respect to the privacy and security of the data subject. For example, this thesis has stated that there should be a minimum to the number of results that are included for the generation of aggregational statistics. While this approach can theoretically prevent information from being exposed, more research is required to determine the threshold in a variety of situations and to determine in general which variables are likely to contribute to jeopardizing the anonymity of the data.

11.3.3 Differences With Traditional Research

If the FAIR-Based Framework discussed in this thesis were to cover Scientific Data Analysis it would create two major differences in the way scientific research is being done compared to traditional research. Section 11.3.3.1 outlines the changes to the accessibility of datasets, specifically what the implications would be of being able to provide access to datasets that were previously unavailable for scientific research. Section 11.3.3.2 outlines the changes to the financial cost of conducting scientific data analysis, specifically the shift in cost from research organization to data collection organization and how funding could be provided under this new system.

11.3.3.1 A Shift in Data Accessibility

If the FAIR-Based Framework discussed in this thesis were to cover Scientific Data Analysis, this would allow a significant amount of data that would normally not be made accessible due to privacy or security concerns to be made available for research. As a result, many more datasets would become available, and given enough measures, a large extent of databases in the healthcare sector could become at least partially accessible for scientific research.

Future work could investigate what the economic value of this would be, which could be used to support the investments required to actually make this system a reality. However, this should also be balanced against the initial findings and future work stated in 11.3.2, which would moderate the value of this approach by reducing the amount of data that would be available and therefore the value that such a change would have.

11.3.3.2 A Shift in Scientific Data Analysis Funding

If the FAIR-Based Framework discussed in this thesis were to cover Scientific Data Analysis, the data visiting approach would shift the funding required to support scientific research from the research organization to the organization that collects the data. This has previously been stated in 11.3.1.2, with the data collection organization now needing additional computer hardware and additional personnel to be able to support this use case.

Future work could investigate the implications of such a shift, specifically, this shift would reduce the level of funding research organizations require for their investigation while data collection organizations would require more funding. However, this raises practical questions related to how this can best be achieved. It would require determining what aspect of scientific research is related to hardware and how much would be saved by data visitation. It would also require determining how the government could best decide to provide these data collection organizations with more funding. It would also require determining how costs could be divided in the fairest way possible, for instance with a computational credit system where organizations would pay for their specific use.

List of Abbreviations and Definitions

Abbreviation	Word	Definition
AcICT	Adviescollege ICT Toetsing	The Adviescollege ICT Toetsing is the permanent replacement organization of the BIT, in charge of evaluating national ICT projects with an IT component of more than 5 million euros
AP	Autoriteit Persoonsgegevens	The Autoriteit Persoonsgegevens is an independent administrative body that has been appointed by law to act as the Dutch data protection authority, in charge of enforcing the GDPR in the Netherlands.
BIO	Baseline informatiebeveiliging Overheid	The Baseline informatiebeveiliging Overheid is a governmental basic standards framework for information security within all levels of government (central government, municipalities, provinces, and the regional water authority). It does not apply to the GGD.
BIT	Bureau ICT-Toetsing	The Bureau ICT-Toetsing was a temporary organization founded in 2015 to improve governmental ICT projects by evaluating ICT projects and providing project-specific recommendations. It was renamed the AcICT in 2020.
BSN	Burgerservicenumber	The Burgerservicenumber is the Dutch term for a citizen service number, which is a numerical value identifying you as a citizen and used as a validation method for sensitive requests
	Data Subject	A data subject is any living individual whose personal data is collected, held, or processed by an organization
CBS	Centraal Bureau voor de Statistiek	The Centraal Bureau voor de Statistiek collects data on Dutch society. This data is processed into statistical information on all kinds of social and economic themes
DPIA	Data Protection Impact Assessment	A Data Protection Impact Assessment is a process resulting in the creation of a document describing the processes used in a project, to identify risks arising out of the processing of personal data and to minimize these risks as far and as early as possible.
EU	European Union	The European Union is a political and economic union of 27 member states, to which the GDPR applies

GDPR	General Data Protection Regulation	The General Data Protection Regulation is the currently acting Data Protection Regulation that went into effect on the 25th of May, 2018
GGD	Gemeentelijke gezondheidsdienst	The Gemeentelijke gezondheidsdienst is a public healthcare organization responsible for protecting, monitoring, and promoting the health of the inhabitants of the Netherlands. The organization itself is federated, consisting of 25 separate GGDs, each in charge of its own region. They each have full autonomy, not even answering to the national government
GGD GHOR	Gemeentelijke gezondheidsdienst en Geneeskundige Hulpverleningsorganisaties in de Regio	GGD GHOR Nederland is the umbrella organization of All GGDs and GHORs in the Netherlands, although this thesis only covers the GGD. This organization is in charge of communication between each of the 25 GGDs, with no authority over any of them.
PVP	Program “Vernieuwd Praeventis”	Program “Vernieuwd Praeventis” refers to the replacement program of Praeventius, the ICT system used to support national vaccination programs
RIVM	Rijksinstituut voor Volksgezondheid en Milieu	The Rijksinstituut voor Volksgezondheid en Milieu is a Dutch research institute in charge of promoting public health, used by the Dutch government to formulate policy
VOG	Verklaring Omtrent het Gedrag	A Verklaring Omtrent het Gedrag is a declaration made by the Dutch government that shows that the judicial past of a person or legal entity does not constitute an objection to the position or purpose for which the VOG has been applied for.
VWS	Ministerie van Volksgezondheid, Welzijn en Sport	The Ministerie van Volksgezondheid, Welzijn en Sport is the Dutch ministry responsible for public health, health care, quality of life, social work, and sport
WOO	Wet open overheid	The Wet open overheid is the successor of the Wet openbaarheid van bestuur, allowing entities to request information about everything the government does

References

[1] I. N. Shu and H. Jahankhani, "The Impact of the new European General Data Protection Regulation (GDPR) on the Information Governance Toolkit in Health and Social Care with Special Reference to Primary Care in England," 2017 Cybersecurity and Cyberforensics Conference (CCC), 2017.

[2] F. Firouzi, B. Farahani, M. Barzegari and M. Daneshmand, "AI-Driven Data Monetization: The Other Face of Data in IoT-Based Smart and Connected Health," *IEEE Internet of Things Journal,* vol. 9, no. 8, pp. 5581-5599, 2022.

[3] W. Wilkowska and M. Ziefle , "Privacy and data security in E-health: requirements from the user's perspective," *Health Informatics Journal,* vol. 18, no. 3, pp. 191-201, 2012.

[4] H. Chen, W. Sun, Z. Cai, Y. Li, F. Liu, S. Fang and G. Wang, "Security and Privacy in the Medical Internet of Things: A Review," *Security and Communication Networks,* vol. 2018, 2018.

[5] F. Schaub, R. Balebako and L. Cranor, "Designing Effective Privacy Notices and Controls," *IEEE Internet Computing,* vol. 21, pp. 70-77, 2017.

[6] OECD, *OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data,* 1980.

[7] Council of Europe, *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108),* Strasbourg , 1981.

[8] European Parliament, *European Parliamentary Research Service,* 2022.

[9] United Kingdom Legislation, *Data Protection Act 1984,* legislation.gov.uk, 1984.

[10] Official Journal of the European Union, *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data,* Official Journal of the European Union, 1995.

[11] United Kingdom Legislation, *Data Protection Act 1998,* 1998.

[12] M. Brodin, "A Framework for GDPR Compliance for Small- and Medium-Sized Enterprises," *European Journal for Security Research,* vol. 4, no. 2, pp. 243-264, 2019.

[13] M. Brodin, "Combining ISMS with strategic management: the case of BYOD," in *8th IADIS International Conference on Information Systems*, Madeira, 2015.

[14] I. Shu and H. Jahankhani, "he Impact of the new European General Data Protection Regulation (GDPR) on the Information Governance Toolkit in Health and Social Care with Special Reference to Primary Care in England," in *Cybersecurity and Cyberforensics Conference*, London, 2017.

[15] C. D. Manatunga, "The GDPR and the use of private permissioned blockchain technology in healthcare sector-Is it possible to comply with the rights of the data subjects? (Master's thesis)," University of Oslo, 2022.

[16] A. Hasselgren, P. Kengfai Wan, M. Horn, K. Kralevska, D. Gligoroski and A. Faxvaag, "GDPR Compliance for Blockchain Applications in Healthcare," NTNU, Norwegian University of Science and Technology, 2020.

[17] D. Tapscott and A. Tapscott, Blockchain revolution: how the technology behind bitcoin and other cryptocurrencies is changing the world, 2018.

[18] C. D. Manatunga, The GDPR and the use of private permissioned blockchain technology in healthcare sector-Is it possible to comply with the rights of the data subjects?, 2022.

[19] A. Dijk, "Success and failure factors in ICT projects: a Dutch perspective.," School of Engineering and Information Sciences Middlesex University, 2009.

[20] A. Payne, "80% of major government projects are at 'risk of failure' as civil servants," *BusinessInsider,* p. 2, 25 January 2018.

[21] CMS, "GDPR Enforcement Tracker - list of GDPR fines," 27 June 2022. [Online]. Available: https://www.enforcementtracker.com/.

[22] H. d. Jonge, "Informatie- en Communicatietechnologie (ICT) in de Zorg," 9 February 2021. [Online]. Available: https://zoek.officielebekendmakingen.nl/kst-27529-235.html.

[23] Regiegroep DOTT, "De Basis op Orde: Verbeterplan fase 1 met betrekking tot Digitale Ondersteuning Test- en Traceerketens," 15 February 2021. [Online]. Available: https://www.rijksoverheid.nl/documenten/publicaties/2021/02/23/basis-op-orde-verbeterplan-dott-210215-10b.

[24] Autoriteit Persoonsgegevens, "Privacy & corona," 9 November 2021. [Online]. Available: https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/corona/privacy-corona.

[25] Stichting ICAM, "Jouw privégegevens op straat door toedoen van de overheid?," 2021. [Online]. Available: https://datalek-ggd.nl/.

[26] L. Van der Plas, "AVG/GDPR Schendingen bij de Coronasystemen van GGD GHOR," 2021.

[27] C. Tankard, "What the GDPR means for businesses," *Network Security,* vol. 6, pp. 5-8, 2016.

[28] The World Bank, *Individuals using the Internet (% of population) - European Union.*

[29] A. Guterres, *Address of the UN Secretary-General to the Italian Senate,* Rome, 2019.

[30] J. P. Albrecht, "How the GDPR Will Change the World," *European Data Protection Law Review,* vol. 2, no. 3, p. 288, 2016.

[31] Official Journal of the European Union, "Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with

regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC," *Official Journal of the European Union,* pp. 1-88, 4 May 2016.

[32] European Commission, "Data protection as a pillar of citizens' empowerment and the EU's approach to the digital transition - two years of application of the General Data Protection Regulation," European Commission, Brussels, 2020.

[33] Official Journal of the European Union, "Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)," *Official Journal of the European Union,* pp. 1-27, 19 December 2002.

[34] Official Journal of the European Council, "Commission Recommendation of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises (Text with EEA relevance) (notified under document number C(2003) 1422)," *Official Journal of the European Council,* p. 1, 2020 May 2003.

[35] GGD GHOR Nederland, "Samen voor een gezond en veilig nederland Jaarbeeld 2020," GGD GHOR Nederland, 2020.

[36] COMMISSION STAFF WORKING DOCUMENT, "COMMISSION STAFF WORKING DOCUMENT Accompanying the document COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL Data protection rules as a pillar of citizens empowerment and EUs approach to digital transition - two years of applic," *Official Journal of the European Council,* 24 June 2020.

[37] KPMG, "Groei AP noodzakelijk voor bescherming burgers: Meerjarenbegroting AP: Een korte toelichting," Autoriteit Perssonsgevens, 2020.

[38] Tweede Kamer der Staten-Generaal, *Motie van het lid Omtzigt over een onderzoek naar de oorzaken van oversterfte,* 2021.

[39] ZonMw , *Beperkt toegang tot data bij onderzoek naar oversterfte,* 2022.

[40] Anoynmous Data Scientist from the Ministry JS, Interviewee, *Correspondence with a Data Scientist from the ministry of Justice and Security.* [Interview]. 30 November 2022.

[41] J. Hansen, P. Wilson, E. Verhoeven, M. Kroneman, M. Kirwan, R. Verheij and E. B. Van Veen, *Assessment of the EU Member States' rules on health data in the light of GDPR,* Brussels, 2021.

[42] M. de Visser, A. Boot, G. Werner, A. van Riel and M. Gijsberts, *Kiezen voor houdbare zorg. Mensen, middelen en maatschappelijk draagvlak WRR-Rapport 104,* 2021.

[43] K. Jones , G. Laurie , L. Stevens , C. Dobbs , D. Ford and N. Lea , *The other side of the coin: Harm due to the non-use of health-related data,* 2017, pp. 43-51.

[44] M. Wilkinson, M. Dumontier, I. Aalbersberg and et al, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data,* vol. 3, pp. 2052-4463, 15 March 2016.

[45] https://fair-dom.org/about, "About FAIRDOM," 2022.

[46] FAIRDOM, "DataHub," FAIRDOM, [Online]. Available: https://fair-dom.org/Datahub. [Accessed 15 January 2023].

[47] FAIRDOM, "IBISBAHub," FAIRDOM, [Online]. Available: https://fair-dom.org/ibisbahub. [Accessed 15 January 2023].

[48] FAIRDOM, "LiSyM," FAIRDOM, [Online]. Available: https://fair-dom.org/LiSyM. [Accessed 1 January 2023].

[49] FAIRDOM , "Leipzig Health Atlas," FAIRDOM , [Online]. Available: https://fair-dom.org/Leipzig. [Accessed 15 January 2023].

[50] FAIRDOM, "NFDI4Health Study Hub Covid-19," FAIRDOM, [Online]. Available: https://fair-dom.org/NFDI4. [Accessed 15 January 2023].

[51] LUMC, "VODAN-Africa researchers design a FAIR digital data health infrastructure that is helping in the fight against the COVID-19 pandemic," LUMC, Leiden, 2021.

[52] Forum Standaardisatie, "Onderzoek naar de stimulering van FAIR Principes bij de overheid," Forum Standaardisatie, 2020.

[53] A. Jacobsen, R. de Miranda Azevedo, N. Juty, D. Batista, S. Coles, R. Cornet, M. Courtot, M. Crosas, M. Dumontier, C. T. Evelo, C. Goble, G. Guizzardi, K. K. Hansen, A. Hasnain and K. Hettne, "FAIR Principles: Interpretations and Implementation," *Data Intelligence,* vol. 2, no. 1-2, pp. 10-29, January 2020.

[54] R. F. van der Lans, Data Virtualization for Business Intelligence Systems, Elsevier, 2012.

[55] O. Osigwe, "Historic First Data Visiting of Patient Data Stored in Residence Across Continents," VODAN AFRICA, 2020.

[56] J. Van Soest, C. Sun, O. Mussmann, M. Puts, B. van den Berg, A. Malic and M. Dumontier, "Building Continents of Knowledge in Oceans of Data: The Future of Co-Created eHealth," *IOS Press,* pp. 581-585, 2018.

[57] GO FAIR, "Personal Health Train," GO FAIR, 2020. [Online]. Available: https://www.go-fair.org/implementation-networks/overview/personal-health-train/. [Accessed 15 January 2023].

[58] Health RI, "Personal Health Train," Health RI, 2018. [Online]. Available: https://www.health-ri.nl/initiatives/personal-health-train. [Accessed 15 January 2023].

[59] S. Y. Amare, G. T. Taye and T. G. Gebreslassie, "Realizing Health Data Interoperability in Low Connectivity Settings: The Case of VODAN-Africa," Unpublished.

[60] W3C, "RDF/XML Syntax Specification (Revised)," W3C, 4 February 2004. [Online]. Available: https://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/. [Accessed 15 January 2023].

[61] M. van Reisen, S. Yohannes, G. Tadele, R. Plug, T. Gebremeskel, A. Aktau and P. H. P. Jati, *FAIR-OLR-Based Digital Health Data,* FAIR Connect, 2023 Forthcoming.

[62] Panteia, "Versterking ICT-werkgeverschap," Panteia, Zoetermeer, 2017.

[63] T. Elias, P. Ulenbelt, M. Fokke, H. B. Slot and P. van Meenen, "Parlementair onderzoek naar ICT-projecten bij de overheid," Tweede Kamer der Staten-Generaal, 2014.

[64] Minister voor Wonen en Rijksdienst, "Instellingsbesluit tijdelijk Bureau ICT-toetsing," Minister voor Wonen en Rijksdienst, 2018.

[65] "Lijst van Zelfstandige bestuursorganen," 2022.

[66] Toezichtsraad Bureau ICT-Toetsing , "Eindrapport: Evaluatie Bureau ICT-Toetsing," Toezichtsraad Bureau ICT-Toetsing , 2018.

[67] D. Stokmans and M. L. Adriaanse, "Ministerie frustreert ict-controle," NRC, 2019.

[68] D. Stokmans and M. L. Adriaanse, "Ministerie vindt toezichthouder ict te kritisch," 2019.

[69] Minister van Justitie en Veiligheid, "Eindrapportage transitie-adviseur relatie onderzoek en beleid op het ministerie van Justitie en Veiligheid," 2021.

[70] Staatssecretaris van Binnenlandse Zaken, "BRIEF VAN DE STAATSSECRETARIS VAN BINNENLANDSE ZAKEN EN KONINKRIJKSRELATIES," Binnenlandse Zaken en Koninkrijksrelaties, 2019.

[71] Adviescollege ICT-Toetsing, "Ministerraad benoemt twee nieuwe leden Adviescollege ICT-toetsing," Adviescollege ICT-Toetsing, 2022.

[72] Adviescollege ICT-Toetsing, "Ministerraad benoemt twee nieuwe leden Adviescollege ICT-toetsing (per 1 januari 2023)," 2022.

[73] Anonymous AcICT employee, Interviewee, *Anonymous interview with an employee of the AcICT.* [Interview]. December 2022.

[74] Adviescollege ICT-Toetsing, "Collegeleden," Adviescollege ICT-Toetsing, 2022.

[75] Adviescollege ICT-toetsing, "Jaarrapportage 2021 - Adviescollege ICT-Toetsing," Adviescollege ICT-toetsing, 2021.

[76] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Rapportage Grote ICT-projecten 2021," Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2021.

[77] A. Van Huffelen, Interviewee, *Interview with the secretary of state of digitalization and kingdomrelations.* [Interview]. 14 September 2022.

[78] Adviescollege ICT-Toetsing, *BIT-advies Beslag Informatie Systeem,* Adviescollege ICT-Toetsing, 2022.

[79] Adviescollege ICT-Toetsing, *Definitief BIT-advies Verbeteren Uitwisseling,* Adviescollege ICT-Toetsing, 2022.

[80] Adviescollege ICT-Toetsing, *Definitief BIT-advies Platform Open OverheidsInformatie,* Adviescollege ICT-Toetsing, 2022.

[81] Adviescollge ICT-Toetsing, *Definitief BIT-advies programma Procesvernieuwing, Informatie,* Adviescollge ICT-Toetsing, 2019.

[82] F. Hendrickx and E. Verwiel, "VWS blijft dwangsommen betalen voor niet openbaren Sywert-communicatie," Volkskrant, 2022.

[83] S. Euwema, "Personal communcation with the Autoriteit Persoonsgevens," Autoriteit Persoonsgevens, 2021.

[84] European Commission, *Europe's Digital Decade: digital targets for 2030,* 2021.

[85] M. Hijink, *Motie van het lid Hijink c.s. over verhoging van het budget van de Autoriteit Persoonsgegevens,* Tweede Kamer Der Staten-Generaal, 2021.

[86] J. Klaver and E. M. J. Ploumen, *Motie van de leden Klaver en Ploumen over uitspreken dat de rechtspositie van burgers versterkt moet worden,* Tweede Kamer Der Staten-Generaal, 2021.

[87] Ministerie van Justitie en Veiligheid, *Vaststelling van de begrotingsstaten van het Ministerie van Justitie en Veiligheid (VI) voor het jaar 2022,* Tweede Kamer der Staten-Generaal, 2021.

[88] NOREA, "NOREA Handreiking Data Protection Impact Assessment," 2020.

[89] Anonymous Data Protection Officer at a ministry, Interviewee, *Anonymous interview with a Data protection Officer at one of the Dutch ministries.* [Interview]. 15 November 2022.

[90] NOREA, "DPIA Raamwerk," NOREA, 2020.

[91] GGD GHOR Nederland, *Data Protection Impact Assessment (DPIA) CoronIT,* 2022.

[92] GGD GHOR Nederland, *Data Protection Impact Assessment (DPIA) GGD Contact ter ondersteuning BCO,* 2022.

[93] Autoriteit Persoonsgegevens, "Overzichten datalekken," 2021.

[94] Autoriteit Persoonsgegevens, "Datalekkenrapportage 2021," 2021.

[95] KPMG, *Onderzoek taken en financiële middelen bij AP,* KPMG, 2020.

[96] Autoriteit Persoonsgegevens, "Meldplicht datalekken: facts & figures Overzicht feiten en cijfers 2020," 2020.

[97] Autoriteit Persoonsgegevens , "Meldplicht datalekken: facts & figures Overzicht feiten en cijfers 2019," 2019.

[98] Autoriteit Persoonsgegevens, "Meldplicht datalekken: facts & figures Overzicht feiten en cijfers 2018," 2018.

[99] Autoriteit Persoonsgegevens, "Datalekkenrapportage 2021," Autoriteit Persoonsgegevens, 2021.

[100] Raad van State, "Wet publieke gezondheid," Raad van State, 2008.

[101] GGD GHOR Nederland, "Jaarbeeld 2021 - Presteren in een turbulente tijd," GGD GHOR Nederland, 2022.

[102] Anonymous GGD management level employee #1 and Anonymous GGD management level employee #2, Interviewees, *Anonymous Interview with two management level employees at one of the GGDs about the early stages of the Covid-19 pandemic and their general experience with governmental administration.* [Interview]. 13 December 2021.

[103] GGD GHOR Nederland, *Personal Communication with a representative of GGD GHOR Nederland,* GGD GHOR Nederland, 2022.

[104] P. Blokhuis, "Nr. 593 BRIEF VAN DE STAATSSECRETARIS VAN VOLKSGEZONDHEID, WELZIJN EN SPORT," *Tweede Kamer der Staten-Generaal,* 20 February 2019.

[105] Rijksinstituut voor Volksgezondheid en Milieu, "Vaccination schedule English: Which vaccines will my child receive?," *Rijksinstituut voor Volksgezondheid en Milieu,* 2021.

[106] C. Hilhorst, *Definitief BIT-advies programma 'Vernieuwd Praeventis,* 2019.

[107] Tweede Kamer der Staten-Generaal, *Jaarverslag en slotwet Ministerie van Volksgezondheid, Welzijn en Sport 2018,* 2019, p. 141.

[108] Ministerie van Volksgezondheid, Welzijn en Sport, *Source Code for the Corona Melder App,* Ministerie van Volksgezondheid, Welzijn en Sport, 2020.

[109] Adviescollege ICT-toetsing, *Evaluatie Ontwikkelproces CoronaMelder App,* 2022.

[110] Ministerie van Volksgezondheid, Welzijn en Sport, *Samenvatting inzichten - onderzoek met gebruikers,* Ministerie van Volksgezondheid, Welzijn en Sport, 2020.

[111] Algemene Bestuursdienst, *CoronaMelder,* 2020.

[112] TenderNed, *Marktconsultatie: Uitnodiging slimme digitale oplossingen Corona,* 2020.

[113] Rijksoverheid, "Opdracht aan GGD-GHOR voor het ontwikkelen en implementeren van CoronIT," Rijksoverheid, 2021.

[114] Raad van State, *Aanbestedingswet 2012,* Raad van State, 2012.

[115] R. Roozendaal, Interviewee, *Interview with the former CIO of the ministry of VWS and the current Deputy DIrector General on Digitalisation.* [Interview]. 19 October 2022.

[116] Allied HR, "2012 Allied Workforce Mobility Survey: Onboarding and Retention," May 2012. [Online]. Available: https://www.allied.com/docs/default-source/pdf/alliedworkforcemobilitysurvey.pdf. [Accessed 2023 January 16].

[117] Tweede Kamer der Staten-Generaal, "Nr. 684 BRIEF VAN DE STAATSSECRETARIS VAN BINNENLANDSE ZAKEN EN KONINKRIJKSRELATIES," *Tweede Kamer der Staten-Generaal,* 29 May 2020.

[118] Tweede Kamer Der Staten Generaal, *Debat over de berichtgeving dat vaccinatie tegen het coronavirus niet meer in 2020 start,* 2020.

[119] Rijksoverheid, "Eerste vaccinatie op 8 januari 2021," Rijksoverheid, 8 January 2021. [Online]. Available: https://www.rijksoverheid.nl/actueel/nieuws/2020/12/17/eerste-vaccinatie-op-8-januari-2021. [Accessed 17 January 2023].

[120] C. Kara-Zaïtri, *Corona-software GGD is niet geschikt voor gebruik tijdens een pandemie,* 2020.

[121] J. van den Berg and C. Huisman, "Met dit houtje-touwtje-systeem gaven GGD-medewerkers de coronacijfers door − tot het crashte," 2020.

[122] S. Van Dam, "Surveillance op de GGD Voorbeeld: HP Zone Dashboard," *GGD Hart voor Brabant.*

[123] Rijksinstituut voor Volksgezondheid en Milieu, *Artikel 26-meldingen Wpg-instellingen,* 2008.

[124] GGD Hart voor Brabant, "Wob-verzoek ICAM – datalek corona 2021," 2021.

[125] GGD GHOR, *Veel gestelde vragen en antwoorden over datadiefstal,* 2022.

[126] RTL Nieuws, *GGD komt met app voor sneller bron- en contactonderzoek,* 2020.

[127] GGD GHOR, *Stand van zaken ontwikkeling digitale ondersteuning bron- en contactonderzoek GGD,* 2021.

[128] H. De Jonge, *Kamerbrief over stand van zaken digitale ondersteuning pandemiebestrijding,* 2021.

[129] J. T. Van Dissel, *Reactie van het RIVM over de impact van overgang naar een nieuw ICT-systeem door het debat over het privacy lek bij de GGD.,* 2021.

[130] J. W. Pauw, Interviewee, *Personal communication with the AcICT.* [Interview]. 11 November 2021.

[131] Tweede Kamer der Staten-Generaal, "Brief van de minister van volksgezondheid, welzijn en sport," 2022.

[132] Staatscourant, "Besluit inzake lijst van verwerkingen van persoonsgegevens waarvoor een gegevensbeschermingseffectbeoordeling (DPIA) verplicht is, Autoriteit Persoonsgegevens," 2019.

[133] Anonymous Data Conference attendee legal, Interviewee, *Anonymous Interview with a legal advisor.* [Interview]. 18 October 2022.

[134] GGD Hart voor Brabant, "Personal Communication with GGD Hart voor Brabant," 2022.

[135] Rijksoverheid, "Gegevensbeschermingseffectbeoordeling (DPIA) COVID-19 notificatie-app," 2020.

[136] Rijksoverheid, "Advies Functionaris voor gegevensbescherming VWS DPIA COVID-19 notificatie-app," 2020.

[137] M. Nieuwenhuis and K. Voskuil, *GGD-medewerkers gluurden ongeoorloofd naar BN'ers in coronadatabase,* AD, 2020.

[138] D. Verlaan, "Privacylek coronasystemen was al maanden bekend, GGD deed niets: 'We konden overal bij'," RTL Nieuws, 2021.

[139] Autoriteit Persoonsgegevens, *Eindbrief onderzoek beveiliging persoonsgegevens GGD GHOR en GGD'en,* 2021.

[140] B. Endedijk and J. Meeus, *Corruptie is bij de Belastingdienst lastig te traceren,* 2022.

[141] M. Nieuwenhuis, *Storing in coronadatabase: gegevens kwijt over kinderen en zorgverleners,* AD, 2020.

[142] J. van Heerde, *De corona-ICT van de GGD wankelt door de drukte. Hoe kan dat?.*

[143] E. Heisen and M. Nieuwenhuis, *Inspectie: Bijna de helft van GGD'en voerde tests op coronavirus niet juist uit,* 2020.

[144] E. Schouten, "Personeelstekorten alom, maar de GGD is uitgegroeid tot megawerkgever," *Nu.nl,* 7 Juli 2021.

[145] H. Hugo, "RTL: Medewerkers waarschuwden sinds zomer over toegang tot coronasystemen GGD," 2021.

[146] F. Damen, *Ik kan zo uitslagen van kennissen zoeken', zeggen medewerkers coronatestlijn,* 2020.

[147] F. Damen, *Testlijnmedewerkers kunnen bij persoonsgegevens, ook als dat niet mag,* 2020.

[148] N. Dekker, *GGD'er Amin (23) lekte gegevens van Peter R. de Vries en John van den Heuvel: 180 uur taakstraf en cel,* 2021.

[149] S. Sietsma and M. Veen, "Persoonsgegevens tienduizenden geteste Nederlanders onvoldoende beveiligd," NOS, 2021.

[150] S. Sietsma and M. Veen, *Inzage in gegevens van tienduizenden geteste Nederlanders, van model tot militair,* 2021.

[151] D. Verlaan, "Illegale handel in privégegevens miljoenen Nederlanders uit coronasystemen GGD," RTL Nieuws, 2021.

[152] NOS, *Eis van 3 miljard om lek testgegevens 'niet heel sympathiek',* 2022.

[153] Nieuwsuur, "Eis van 3 miljard om lek testgegevens 'niet heel sympathiek'," 2021.

[154] Fox-IT, "Vacatures - Werken bij Fox-IT," Fox-IT, 2023. [Online]. Available: https://www.fox-it.com/nl/vacatures/#. [Accessed 19 January 2023].

[155] GGD GHOR, *GGD en haar data – Hoe zit het echt? Een repliek,* 2021.

[156] F. Damen and C. N. Bijvank, *Waakhond stelt GGD onder 'verscherpt toezicht', problemen al maanden bekend,* 2021.

[157] M. J. Verdier , *Onderzoek Beveilig GGD corona,* 2021.

[158] D. Verlaan, *Datadiefstal GGD veel groter dan gemeld, gedupeerden niet geïnformeerd,* 2021.

[159] M. Hommes, *Datalek GGD: schadevergoeding voor 1.250 gedupeerden,* 2022.

[160] Openbaar Ministerie, *Celstraf geëist voor diefstal en verkoop van gegevens GGD,* 2022.

[161] Stichting Initiatieven Collectieve Acties Massaschade, *Veelgstelde Vragen.*

[162] GDPRHUB, "Article 24 - Responsibility of the controller," GDPRHUB, 25 April 2022. [Online]. Available: https://gdprhub.eu/Article_24_GDPR. [Accessed 21 January 2023].

[163] M. Lang, DSGVO - BDSG - TTDSG Art. 24 DS-GVO margin numbers 23-24, 4 ed., 2022.

[164] M. Martini, *DSGVO - BDSG - TTDSG Art. 24 DS-GVO margin numbers 21-22,* 4 ed., 2021.

[165] Privacy Regulation, "Recital 78 EU GDPR," Privacy Regulation, [Online]. Available: https://www.privacy-regulation.eu/en/r78.htm. [Accessed 19 January 2023].

[166] Autoriteit Persoonsgegevens, "AP eist opheldering van GGD.," Autoriteit Persoonsgegevens, 2021.

[167] R. Damhoff, Interviewee, *Interview with the Chief Data Officer of the ministry of Justice and Security.* [Interview]. 18 October 2022.

[168] Digitale Overheid, "Baseline Informatiebeveiliging Overheid," Digitale Overheid, 23 January 2023. [Online]. Available: https://www.digitaleoverheid.nl/overzicht-van-alle-onderwerpen/cybersecurity/kaders-voor-cybersecurity/baseline-informatiebeveiliging-overheid/. [Accessed 23 January 2023].

[169] Nederlandse Overheid Referentie Architectuur, "BIR (Baseline Informatiebeveiliging Rijksdienst)," 2012.

[170] Nederlandse Overheid Referentie Architectuur, "BIG (Baseline Informatiebeveiliging Gemeenten)," 2013.

[171] Centraal Informatie Beveiligings Overleg, "IBI (Interprovinciale Baseline Informatiebeveiliging)," 2010.

[172] Unie van Waterschappen, "Baseline Informatiebeveiliging Waterschappen," 1 October 2013. [Online]. Available: https://www.noraonline.nl/images/noraonline/0/09/Baseline-Informatiebeveiliging-Waterschappen-2013.pdf.

[173] Nederlands Instituut Publieke Veiligheid, *Versnellingsplan Informatieveiligheid,* 2022.

[174] Nederlands Instituut Publieke Veiligheid, "Personal correspondence with the Nederlands Instituut Publieke Veiligheid," 2022.

[175] Rijksoverheid, "Evaluatie Baseline Informatiebeveiliging Overheid (BIO)," Rijksoverheid, 2021.

[176] Normcommissie 303006 ‘Informatievoorziening in de zorg’, "NEN 7513:2018 nl," Normcommissie 303006 ‘Informatievoorziening in de zorg’, 2018.

[177] A. Wokke, *Tientallen medewerkers Haags ziekenhuis bekeken medisch dossier van bn'er,* 2018.

[178] Chipsoft, *HagaZiekenhuis succesvol over op HiX,* Chipsoft.

[179] G. van Gorp, *Ambtenaren verlekkeren zich aan privégegevens BN'ers,* 2014.

[180] Anoynmous iBestuur attendee from the DJI, Interviewee, *Anonymous interview with a representative from the Custodial Institutions Agency.* [Interview]. 14 September 2022.

[181] Anonymous iBestuur attendee from the DIVD, Interviewee, *Anonymous interview with a representative of the Dutch Institute for Vulnerability Disclosure.* [Interview]. 14 September 2022.

[182] Anonymous Data Conference attendee, Interviewee, *Anonymous interview with a representative of the Inspectorate of Justice and Security.* [Interview]. 18 October 2022.

[183] Anonymous iBestuur attendee from the NCTV, Interviewee, *Anonymous interview with a representative of the National Coordinator for Security and Counterterrorism.* [Interview]. 14 September 2022.

[184] Anonymous iBestuur attendee NFI, Interviewee, *Interview with a representative of the Netherlands Forensic Institute.* [Interview]. 14 September 2022.

[185] Anonymous Data Conference attendee from the NFI, Interviewee, *Anonymous interview with a representative of the Netherlands Forensic Institute.* [Interview]. 18 October 2022.

[186] A. Oostenbrug, P. Becker, M. van Wallenburg, L. Kool and M. Fraanje, Interviewees, *https://www.ibestuurcongres.nl/2022/sessies.* [Interview]. 14 September 2022.

[187] Anonymous iBestuur attendee from the DG Digital, Interviewee, *Anonymous interview with a representative of the Department General of Digitalisation.* [Interview]. 14 September 2022.

[188] Anonymous iBestuur attendee from the ministry JS, Interviewee, *Anonymous interview with a representative of the ministry of Justice and Security.* [Interview]. 14 September 2022.

[189] *Data Conference Panel about Data Processing and Sharing,* 2022.

[190] RIVM, "Eenheid van taal in de Nederlandse Zorg," RIVM, 2018.

[191] M. Kedilioglu, Interviewee, *Interview with Mustafa, project manager in the Ministry of the Interior and Kingdom Relations.* [Interview]. 2022.

[192] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Broncode DigiD app," Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 11 January 2023. [Online]. Available: https://github.com/MinBZK/woo-besluit-broncode-digid-app. [Accessed 26 January 2023].

[193] A. C. van Huffelen, *Kamerbrief over openbaarmaking broncode DigiD-app,* 2023.

[194] A. C. van Huffelen, *Wijziging van het voorstel van wet houdende algemene regels inzake het elektronisch verkeer in het publieke domein en inzake de generieke digitale infrastructuur (Wet digitale overheid),* 2022.

[195] "Het Algoritmeregister van de Nederlandse overheid," 21 December 2022. [Online]. Available: https://algoritmes.overheid.nl/. [Accessed 26 January 2023].

[196] A. C. Van Huffelen , *Kamerbrief over het algoritmeregister,* Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 2022.

[197] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "About the Algorithm Register," Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 20 December 2022. [Online]. Available: https://github.com/MinBZK/Algoritmeregister. [Accessed 26 January 2023].

[198] Z-CERT, "Cybersecurity Dreigingsbeeld Zorg 2021," Z-CERT, 2021.

[199] Hall Booth Smith, P.C. Attorneys At Law, "We All Know About GDPR's Right to Erasure, Does This Mean You Have to Delete Data From Backups As Well?," Hall Booth Smith, P.C. Attorneys At Law, [Online]. Available: https://hallboothsmith.com/we-all-know-about-gdprs-right-to-erasure-does-this-mean-you-have-to-delete-data-from-backups-as-well/. [Accessed 25 January 2023].

[200] Autoriteit Persoonsgegevens, "Vallen back-ups ook onder het recht op vergetelheid?," Autoriteit Persoonsgegevens, 2019. [Online]. Available: https://autoriteitpersoonsgegevens.nl/nl/onderwerpen/algemene-informatie-avg/rechten-van-betrokkenen. [Accessed 25 January 2023].

[201] Centric, "Overheidssites onvoldoende bereikbaar," Centric, 16 Oktober 2008. [Online]. Available: https://www.binnenlandsbestuur.nl/digitaal/centric/overheidssites-onvoldoende-bereikbaar. [Accessed 26 January 2023].

[202] Z. Flower, "Why is high availability important in cloud computing?," TechTarget Cloud Computing, 22 October 2020. [Online]. Available: https://www.techtarget.com/searchcloudcomputing/answer/How-much-cloud-uptime-do-you-need. [Accessed 25 January 2023].

[203] Bundesamt für Sicherheit in der Informationstechnik, "Tipps für sicheres mobiles," Bundesamt für Sicherheit in der Informationstechnik, 14 April 2021. [Online]. Available: https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Cyber-Sicherheit/Themen/empfehlung_home_office.pdf?__blob=publicationFile&v=4. [Accessed 25 April 2021].

[204] European Commission, *Recital 89 EU GDPR,* European Commission, 2018.

[205] Kennis- en Exploitatiecentrum voor Officiële Overheidspublicaties, "TOOI - Inleiding 1.0.0," Standaarden Overheid, 6 December 2022. [Online]. Available: https://standaarden.overheid.nl/tooi/doc/tooi-inleiding/.

[206] Standaarden Overheid, "TOOI," Standaarden Overheid, [Online]. Available: https://standaarden.overheid.nl/tooi. [Accessed 10 Juli 2023].

[207] Kennis- en Exploitatiecentrum voor Officiële Overheidspublicaties, "TOOI - Ontologie 1.1.2," Standaarden Overheid, 29 June 2023. [Online]. Available: https://standaarden.overheid.nl/tooi/doc/tooi-ontologie/. [Accessed 10 Juli 2023].

[208] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Register van Overheidsorganisaties," Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, [Online]. Available: https://organisaties.overheid.nl/. [Accessed 10 July 2023].

[209] Gemeente Katwijk, "Gemeente Katwijk (Katwijk)," Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 8 July 2023. [Online]. Available: https://organisaties.overheid.nl/27030/Gemeente_Katwijk. [Accessed 10 July 2023].

[210] Provincie Zuid-Holland, "Provincie Zuid-Holland (ZH)," Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, 1 September 2023. [Online]. Available: https://organisaties.overheid.nl/17466/Provincie_Zuid-Holland. [Accessed 10 July 2023].

[211] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Jaarrapportage Bedrijfsvoering Rijk 2019 Bijlage 3 Grote ICT-projecten," 2020.

[212] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Jaarrapportage Bedrijfsvoering Rijk 2020 Bijlage 3 Grote ICT-projecten," 2021.

[213] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Jaarrapportage Bedrijfsvoering Rijk 2021 Bijlage 3 Grote ICT-projecten," 2022.

[214] Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, "Jaarrapportage Bedrijfsvoering Rijk 2022 Bijlage 3 Grote ICT-projecten," 2023.

[215] Rijksoverheid, "Rijks ICT-dashboard," Rijksoverheid, 13 July 2023. [Online]. Available: https://rijksictdashboard.nl/. [Accessed 13 July 2023].

[216] Rijksoverheid, "Dataregister van de Nederlandse Overheid," 2022.

[217] Nictiz, "SNOMED CT," Nationaal ICT-Instituut in de Zorg, 6 April 2018. [Online]. Available: https://nictiz.nl/standaarden/overzicht-van-standaarden/snomed-ct/. [Accessed 2023 July 19].

[218] Nictiz, "Belangrijke stap in digitale gegevensuitwisseling tussen huisarts en ziekenhuis," Nationaal ICT-Instituut in de Zorg, 7 October 2021. [Online]. Available: https://nictiz.nl/nieuws/belangrijke-stap-in-digitale-gegevensuitwisseling-tussen-huisarts-en-ziekenhuis/.

[219] Nictiz, "Vertalen SNOMED," Nationaal ICT-Instituut in de Zorg, 2017. [Online]. Available: https://nictiz.nl/wat-we-doen/programmas/archief/vertalen-snomed/. [Accessed 19 July 2023].

[220] ICT&health, "SNOMED: MEESTE MEDISCHE TERMEN IN NEDERLANDS BESCHIKBAAR," ICT&health, 7 October 2023. [Online]. Available: https://icthealth.nl/nieuws/snomed-meeste-medische-termen-in-nederlands-beschikbaar/.

[221] Rijksoverheid, "Hoe kom ik aan een burgerservicenummer (BSN)?," Rijksoverheid, [Online]. Available: https://www.rijksoverheid.nl/onderwerpen/privacy-en-persoonsgegevens/vraag-en-antwoord/hoe-kom-ik-aan-een-burgerservicenummer-bsn. [Accessed 17 July 2023].

[222] Rijksoverheid, "Basisregistratie Personen (BRP)," Rijksoverheid, [Online]. Available: https://www.rijksoverheid.nl/onderwerpen/privacy-en-persoonsgegevens/basisregistratie-personen-brp. [Accessed 17 July 2023].

[223] D. Brickley and L. Miller, "FOAF Vocabulary Specification," 18 August 2004. [Online]. Available: http://xmlns.com/foaf/0.1/. [Accessed 17 July 2023].

[224] Rijksoverheid, "Gedeelde persoonsgegevens," Rijksoverheid, [Online]. Available: https://mijn.overheid.nl/identiteit/gedeelde-persoonsgegevens. [Accessed 17 July 2023].

[225] Logius, "Tarieven," Logius, [Online]. Available: https://www.logius.nl/onze-organisatie/zakendoen-met-logius/doorbelasting#. [Accessed 01 April 2023].

[226] Logius, "Eerste viermaandsrapportage Logius 2020," Logius, 2020. [Online]. Available: https://www.logius.nl/onze-organisatie/logius-rapporteert/eerste-viermaandsrapportage-logius-2020. [Accessed April 01 2023].

[227] Rijksoverheid, "Over dit dashboard," Rijksoverheid, [Online]. Available: coronadashboard.rijksoverheid.nl/over. [Accessed 11 March 2023].

[228] Rijksoverheid, "View on patients," Rijksoverheid, 9 March 2023. [Online]. Available: https://coronadashboard.government.nl/landelijk/patienten-in-beeld.

[229] Rijksinstituut voor Volksgezondheid en Milieu, "COVID-19 dataset," Rijksinstituut voor Volksgezondheid en Milieu, [Online]. Available: https://data.rivm.nl/covid-19/. [Accessed 14 March 2023].

[230] Overheid, *Wet aanvullende bepalingen verwerking persoonsgegevens in de zorg,* Overheid, 2020.

[231] Logius , "Hoe werkt DigiD?," Logius, [Online]. Available: https://logius.nl/domeinen/toegang/digid/hoe-werkt-het. [Accessed 30 January 2023].

[232] OpenAI, "GPT-4 is OpenAI's most advanced system, producing safer and more useful responses," OpenAI, 14 March 2023. [Online]. Available: https://openai.com/gpt-4. [Accessed 22 July 2023].

[233] A. Panagopoulos†, T. Minssen, K. Sideri, H. Yu and M. C. Compagnucci, "Incentivizing the Sharing of Healthcare Data in the AI Era," *Computer Law & Security Review,* vol. 45, p. 105670, 2022.

[234] M. Gaikema, M. Donkersloot, J. Johnson and H. Mulder, "Increase the success of Governmental IT-projects," CHAOS University System, 2019.

[235] DLA Piper, "DLA Piper GDPR fines and data breach survey: January 2022," DLA Piper, 2022.

[236] GDPR.eu, "General Data Protection Regulation (GDPR) Compliance Guidelines," 19 February 2019. [Online]. Available: https://gdpr.eu/.

[237] U. Mustafa, E. Pflugel and N. Philip, "A Novel Privacy Framework for Secure M-Health Applications: The Case of the GDPR," *2019 IEEE 12th International Conference on Global Security, Safety and Sustainability (ICGS3),* pp. 1-9, 2019.

[238] M. Taddicken, "The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure," *Journal of Computer-Mediated Communication,* vol. 19, no. 2, pp. 248-273, 2014.

[239] R. Sanders, "Eerste leden voor opvolger van Bureau ICT-toetsing," Computable.nl, 2021.

[240] BNR Webredactie, "Ton Elias over falend ICT-project: Te raar voor woorden," *BNR,* 27 January 2022.

[241] Rijksoverheid, "Veiligheidsregio's," Rijksoverheid, 2019.

[242] European Commision, "For how long can data be kept and is it necessary to update it?," European Commision, 2016.

[243] Autoriteit Persoonsgegevens, "Data protection impact assessment," Autoriteit Persoonsgegevens, 2019.

[244] D. Verlaan, "Privacywaakhond: GGD moet privégegevens burgers beter beschermen," RTL Nieuws, 2021.

[245] H. de Jonge, "Beantwoording Kamervragen over datalek coronasystemen GGD," Ministerie van Volksgezondheid, Welzijn en Sport, 2021.

[246] DOMO, "Data Never Sleeps 5.0," 2018. [Online]. Available: https://www.domo.com/learn/infographic/data-never-sleeps-5.

[247] Global Observatory, "Global Observatory for eHealth," in *Fifty-eighth World Health Assembly*, 2005.

[248] Secretary of state, Ministry of Health, Welness and Sport, *Response to the 'Definitief BIT advies programma "Vernieuwd Praeventis"',* 2019.

[249] RIVM, *Dutch National Immunisation Programme.*

[250] K. Scarfone, W. Jansen and M. Tracy, *Recommendations of the National Institute of Standards and Technology - Guide to General Server,* 2008.

[251] Rijksoverheid, *Nu ook mogelijk om online afspraak coronatest te maken,* 2020.

[252] *GGD-directeur: 'Als de kraan steeds verder open wordt gezet, is geen enkele dweil groot genoeg',* 2020.

[253] H. d. Jonge, *Kamerbrief over COVID-19 vaccinatiestrategie update stand van zaken,* 2021.

[254] W. Enserink, *Corona App Hackathon: lof of kritiek?,* 2020.

[255] Rijksoverheid, *Handboek portfoliomanagement,* 2014.

[256] A. C. van Huffelen, M. A. Adriaansens and D. Yeşilgöz-Zegerius, *Kamerbrief hoofdlijnen beleid voor digitalisering.*

[257] Nederlandse Overheid Referentie Architectuur, *RFC Bindende architectuurafspraken 2022,* 2022.

[258] Nederlandse Overheid Referentie Architectuur, *Principes,* 2011.

[259] Nederlandse Overheid Referentie Architectuur, *Publieke Gezondheid Referentie Architectuur,* 2019.

[260] Digitale Overheid, *Stelselplaat gegevens en basisregistraties,* 2020.

[261] Digitale Overheid, *I-strategie Rijk 2021 - 2025 Doorpakken op Digitale Transformatie,* 2021.

[262] GGD GHOR, *Q&A WOB-/WOO-VERZOEK,* 2021.

[263] M. van de Klundert and J. Schellevis, "Lek in GGD-systeem al driekwart jaar aanwezig," NOS.

[264] Information Commissioner's Office, "What does it mean if you are joint controllers?," Information Commissioner's Office, 2018.

[265] M. Van Reisen, F. Oladipo, M. Stokmans, M. Mpezamihgo, S. Folorunso, E. Schultes, M. Basajja, A. Aktau, S. Y. Amare, G. T. Taye, P. Hadi , P. Jati, K. Chindoza, M. Wirtz, M. Ghardallou, G. v. S. Van Stam, W. Ayele, R. Nalugala, I. Abdullahi, O. Osigwe, J. Graybeal, A. A. Medhanyie, A. A. Kawu, F. Liu, K. Wolstencroft, E. Flikkenschild, Y. Lin, J. Stocker and M. A. Musen, *Design of a FAIR digital data health infrastructure in Africa for COVID-19 reporting and research,* vol. 2, 2021.

[266] Nictiz, *Nederlandse kennisorganisatie voor digitale informatievoorziening in de zorg.*

[267] E. Meijer, *Fox-IT gaat logs van GGD-systeem CoronIT continu controleren,* 2021.

[268] Kaggle, "Titanic - Machine Learning from Disaster," Kaggle, [Online]. Available: https://www.kaggle.com/c/titanic/data. [Accessed 27 February 2023].

[269] Centraal Bureau voor de Statistiek, "Regionale kerncijfers Nederland," Centraal Bureau voor de Statistiek, [Online]. Available: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/70072NED/table?fromstatweb. [Accessed 15 March 2023].

Appendix

A1. WOO-verzoek updated DPIA CoronIT related to the vaccination functionalities added at a later stage (2022.263)

Beste heer/mevrouw,

Mijn naam is Leendert van der Plas, telefoonnummer: [redacted] - [redacted] en emailadress: [redacted], adres: [redacted], [redacted] [redacted].

Bij deze doe ik een Woo-verzoek naar de DPIA die rond de periode **Juni-December 2020** gemaakt moet zijn door GGD GHOR voor CoronIT voor het verwerken van medische gegevens omtrent vaccinaties. Waarbij het Ministerie van Volksgezondheid, Welzijn en Sport de opdrachtgever is en GGD GHOR Nederland de opdrachtnemer en het gebruikt wordt in alle GGDen.

Hier is reeds op 15 februari 2022 een verzoek voor gedaan bij door stichting ICAM. Echter is deze informatie niet compleet. Er is inderdaad een DPIA van CoronIT gepubliceerd, maar het lijkt hier om een eerdere versie te gaan. In het document 'https://www.ggdzeeland.nl/app/uploads/2022/06/categorie-viii-8.2-20210326_DPIA-1-juni-2022.pdf', wordt niet gesproken over het vaccinatie aspect van CoronIT. Welke een latere toevoeging is geweest op CoronIT, aangezien vaccineren pas vanaf Januari 2021 nodig was.

Hieruit concludeer ik dat het hier om de initiele DPIA gaat die opgesteld is voor het verwerken van medische informatie omtrent het test process. Graag ontvang ik de tweede DPIA van CoronIT die gemaakt moet zijn voor het verwerken van medische informatie omtrent het vaccinatie process.

In het geval dat dit document niet bestaat, zowel in een compleet nieuwe DPIA omtrent het verwerken van medische informtie bij het vaccinatie process, of een herziene DPIA waarin beshreven staat hoe medische informatie wordt het ondersteunen van het test EN vaccinatie proces, wil ik hier graag bevestiging van.

Met vriendelijke Groet,
Leendert van der Plas

A2. WOO-verzoek DPIA HPZone and HPZone Lite (2022.264)

Beste heer/mevrouw,

Mijn naam is Leendert van der Plas, telefoonnummer: [redacted] - [redacted] en emailadress: [redacted], adres: [redacted], [redacted] [redacted].

Bij deze doe ik een Woo-verzoek naar de DPIA die rond de periode **Maart-Augustus 2020** gemaakt moet zijn door GGD GHOR voor HPZone Lite, de herziene versie van HPZone. Aangezien het hier gaat om een nieuw systeem, of een systeem met een aanzielijk grotere schaal en met een andere type medewerkers dan medische professionals moet hier een DPIA voor gemaakt zijn. Graag ontvang ik deze DPIA. Als dit document niet bestaat, wil ik hier graag een bevestiging van.

Ook ontvang ik graag de DPIA die gemaakt is voor de uitbreiding van HPZone, voor de periode tussen Maart en Augusts 2020. Gedurende deze periode is het orginele systeem, althans voor dezelfde doeleiden ingezet op een aanzienlijk grotere schaal. Als dit document niet bestaat, wil ik hier graag een bevestiging van.

Deze systemen zijn in gebruik bij alle GGDen, uitgevoerd door GGD GHOR Nederland. Met als eindverantwoordelijke het Ministerie van Volksgezondheid, Welzijn en Sport.

Met Vriendelijke Groet,
Leendert van der Plas

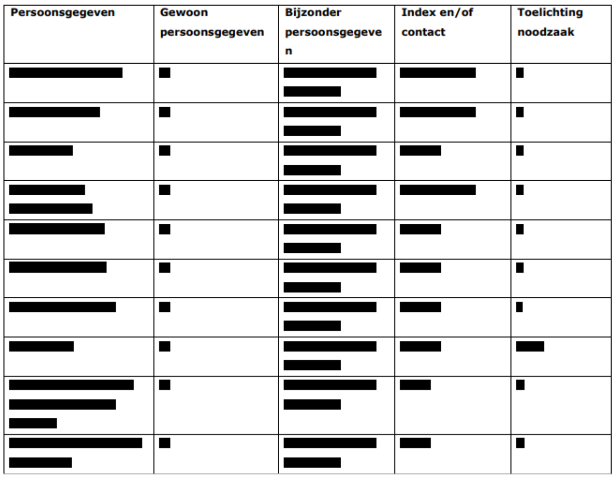
A3. WOO-verzoek DPIA GGD Contact (2022.265)

Beste heer/mevrouw,

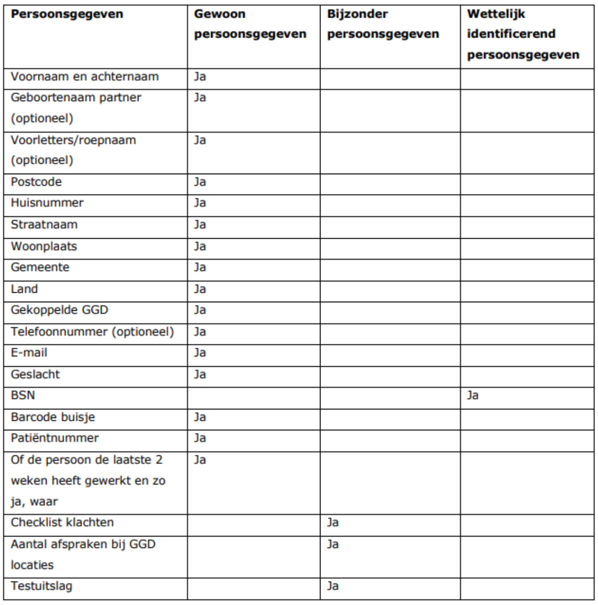
Mijn naam is Leendert van der Plas, telefoonnummer: [redacted] - [redacted] en emailadress: [redacted], adres: [redacted], [redacted] [redacted].

Bij deze doe ik een Woo-verzoek naar de persoonsgegevens sectie van de DPIA die gemaakt is voor GGD Contact, in de periode **Februari 2021 tot eind 2022**, aangezien het systeem nog steeds niet volledig klaar is. Waarbij het Ministerie van Volksgezondheid, Welzijn en Sport de opdrachtgever is en GGD GHOR Nederland de opdrachtnemer en het gebruikt wordt (of gaat worden) in alle GGDen.

Hier is reeds op 15 februari 2022 een verzoek voor gedaan door stichting ICAM. Maar in dit document is, voor voor mij onduidelijke redenen, ervoor gekozen om de data tabel met gegevens die in het systeem staan zwart gemaakt. Ik heb deze tabel hieronder toegevoegd. Deze DPIA is gepubliceerd en de betreffende tabel staat op pagina 13 van dit document 'https://www.ggdzeeland.nl/app/uploads/2022/06/Categorie-viii-8.1_DPIA_-20210412.pdf'.



Persoonsgegeven	Gewoon persoonsgegeven	Bijzonder persoonsgegeven	Index en/of contact	Toelichting noodzaak
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]
[redacted]	[redacted]	[redacted]	[redacted]	[redacted]



Persoonsgegeven	Gewoon persoonsgegeven	Bijzonder persoonsgegeven	Wettelijk identificerend persoonsgegeven
Voornaam en achternaam	Ja		
Geboortenaam partner (optioneel)	Ja		
Voorletters/roepnaam (optioneel)	Ja		
Postcode	Ja		
Huisnummer	Ja		
Straatnaam	Ja		
Woonplaats	Ja		
Gemeente	Ja		
Land	Ja		
Gekoppelde GGD	Ja		
Telefoonnummer (optioneel)	Ja		
E-mail	Ja		
Geslacht	Ja		
BSN			Ja
Barcode buisje	Ja		
Patiëntnummer	Ja		
Of de persoon de laatste 2 weken heeft gewerkt en zo ja, waar	Ja		
Checklist klachten		Ja	
Aantal afspraken bij GGD locaties		Ja	
Testuitslag		Ja	

Figure 59 – Comparison between the personal data table from the DPIA of GGD Contact (left) and the DPIA of CoronIT (right)

In ditzelfde Woo-verzoek is de data tabel van CoronIT wel openbaar 'https://www.ggdzeeland.nl/app/uploads/2022/06/categorie-viii-8.2-20210326_DPIA-1-juni-2022.pdf'. Ik zie dus niet in wat de grondslag is van het feit dat deze informatie verwijderd is uit de DPIA van GGD Contact. Graag ontvang ik of deze data tabel in zijn volledigheid, of de grondslag waarom het nodig geacht werd om deze informatie niet zichtbaar te maken. En waarom er voor gekozen is om precies dezelfde tabel bij de DPIA can CoronIT wel zichtbaar te maken

Met Vriendelijke Groet,
Leendert van der Plas

A4. Timeline of important events related to the Covid-19 Pandemic and the GGD

Date	Event
2020 January	Gathering of EU leaders after rumors of a deadly virus in China
2020 February 27th	The first case of Covid-19 detected in the Netherlands
2020 March-April	First intelligent lockdown in the Netherlands
2020 March-July	Different GGD regions are all working in different systems
2020 June 1st	Testing has been expanded to allow even people with minor symptoms to be able to go to a testing facility
2020 June 8th	Project Tender signed by GGD GHOR
2020 July 15th	System crashes due to high demand, leading to disruptions of service and a loss of data
2020 Summer	Internal reports in the GGD about privacy risks, no actions are taken
2020 August 12th	Start of being able to make an online appointment for a corona test
	GGD stops large-scale source and contact research
2020 November	Start development and initial test of GGD Contact, an app to aid in source and contact tracing
2020 November 3rd	First notice of a data breach to criminals reported. GGD: “No restrictions necessary”
2020-2021 January	Many complaints by staff. Deemed non-important
2020 Mid-November/December	Start of development of vaccination side of CoronIT
2020 December 31st	Final ‘dry run’ test of the vaccination side of CoronIT
2021 January 8th	First vaccination with the Covid-19 vaccine
2021 January 21st	RTL news reports data breach at U-diagnostic. A private company processing covid-19 related data
2021 January 22nd	GGD GHOR Nederland notifies the AP of a data breach
2021 January 25th	RTL News reports that a massive data breach occurred at the GGD. Data from tens of thousands of people have been sold
2021 January 27th	AP demands clarification concerning the massive data breach reported by RTL Nieuws
2021 February 15th	Audit of CoronIT and HPZone Lite by an external consulting agency identifies major problems and advises on solutions
2021 February	The decision was made to replace HPZone Lite
2021 End of March	Planned release date of automated suspicious behavior identification system
2021 April	The decision to make GGD Contact (System) the replacement for HPZone Lite
2021 September 12th	GGD GHOR reports that 1250 data subjects have been notified, not tens of thousands
2021 November 11th	AP reports that personal data is still not processed securely enough, and threatens with fines. The automated suspicious behavior identification system has still not been implemented
2022 Fall	HPZone Lite for combating COVID-19 is expected to be finally phased out in the fall of 2022, HPZone Lite will then revert to its function before the pandemic

Table 26 - Timeline of important events related to the Covid-19 Pandemic and the GGD

A5. Basisregisters

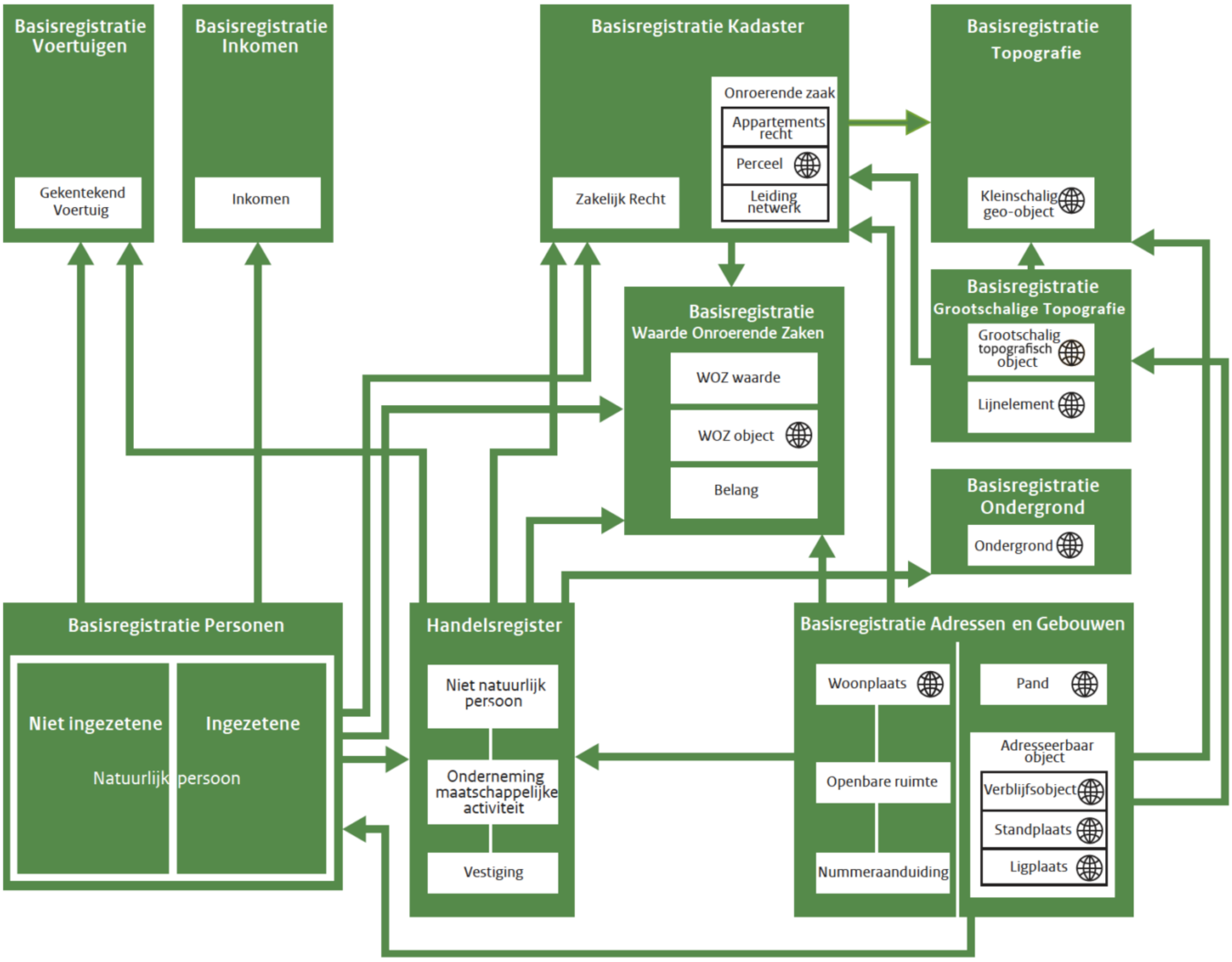


Figure 60 - Overview of the 'Basisregisters' and underlying links