

Security of Big Data in Healthcare Systems

How to ensure security in interoperable healthcare systems with the use of frameworks for security governance and risk management

VEGARD SAAVESEN MATHISEN
ØYVIND HAMMER MARKENG

SUPERVISOR
Paolo Spagnoletti

University of Agder, 2022
Faculty of Social Sciences
Department of Information Systems

PREFACE

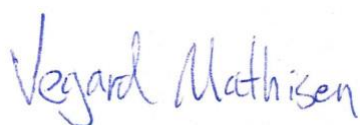
This thesis is the culmination of the cybersecurity master's program at the University of Agder (UiA), in the faculty of social science – department of Information Systems.

We would like to express our gratitude to our supervisor, Paolo Spagnoletti at the Department of Information Systems at the University of Agder, for providing excellent comments, pointing us in the right direction, encouraging and motivating us through in the process of writing this master thesis.

We would also like to thank Eirik Thormodsrud at Sopra Steria/Möller Mobility Group for guidance, supervision, and assistance in finding interview subjects and other useful information related to the thesis.

Lastly, we would like to thank all the interview subjects who took time of their day to participate in our interviews.

Kristiansand,
June 3rd, 2022



Vegard Saavesen Mathisen



Øyvind Hammer Markeng

ABSTRACT

As cyber-attacks have become more common and sophisticated, the need for a stable security framework has become essential. Information security requirements must be met by digital technologies utilized in the health care sector. Modern hospitals are becoming increasingly digital, and information and communication technology is becoming an increasingly significant element of the core business. This lays the groundwork for improved patient care quality. At the same time, the health sector's vulnerability to digital attacks and data breaches is growing, and so are the potential negative effects of security breaches.

The Norwegian healthcare system is divided into different regions, each with its own set of processes and procedures. Because of the fragmentation, there are substantial communication issues between the many health regions and their systems, making transmitted data vulnerable to threat actors. A reorganization is required to effectively handle this issue and improve the security of healthcare systems.

The research was conducted using a qualitative method with a problem-oriented phenomenon-driven research approach on Norwegian Healthcare Sector. In addition, interviews with different security employees from the different health regions in Norway, as well as a document analysis of published papers was done to gather empirical material for the master thesis.

Keywords

Risk Management, Information Security Governance, Interoperability, Big Data, Privacy, Healthcare ICT

Table of contents

PREFACE	III
ABSTRACT	V
LIST OF ABBREVIATION	X
1 INTRODUCTION.....	1
1.1 RESEARCH BACKGROUND.....	1
1.2 RESEARCH PROBLEM.....	4
1.3 RESEARCH QUESTIONS.....	5
1.4 THESIS STRUCTURE	5
2 THEORETICAL BACKGROUND	7
2.1 SELECTED LITERATURE.....	7
2.2 REVIEW FINDINGS.....	9
2.2.1 <i>Electronic Health Records</i>	9
2.2.2 <i>Privacy in Healthcare</i>	11
2.2.3 <i>Big Data Governance</i>	12
2.2.4 <i>Interoperability Framework</i>	15
2.3 DISCUSSION ON THE LITERATURE REVIEW.....	17
3 RESEARCH APPROACH.....	19
3.1 PHENOMENON-DRIVEN RESEARCH.....	19
3.2 LITERATURE REVIEW	20
3.3 QUALITATIVE RESEARCH APPROACH	24
3.4 RESEARCH INTERVIEWS.....	25
3.5 DOCUMENT ANALYSIS	26
3.6 DATA COLLECTION PROCESS.....	27
3.6.1 <i>Selection of Interview Subjects</i>	27
3.6.2 <i>Analyzing Interviews</i>	28
3.6.3 <i>Document Selection</i>	29
3.7 MASTER THESIS DATA COLLECTION PROCESS.....	30
3.8 VALIDITY & LIMITATION.....	31
3.9 ETHICAL CONSIDERATIONS.....	32
4 FINDINGS	33
4.1 BIG DATA IN HEALTHCARE.....	33
4.1.1 <i>Privacy</i>	34

4.1.2	<i>Other Challenges</i>	35
4.2	RISK MANAGEMENT	37
4.3	GOVERNANCE	39
4.4	INTEROPERABILITY.....	43
5	DISCUSSION	46
5.1	PRIVACY.....	46
5.2	RISK MANAGEMENT	47
5.3	INFORMATION SECURITY GOVERNANCE	51
5.4	INTEROPERABILITY FRAMEWORKS	52
5.5	FINAL RESULT	55
6	CONCLUSION	58
7	LIMITATIONS	59
7.1	RESEARCH PROBLEM.....	59
7.2	INTERVIEWS	59
7.3	SAMPLE	60
7.4	TIME CONSTRAINTS.....	60
8	FURTHER RESEARCH	61
	RESOURCES	62
	APPENDIX A	67
	INTERVIEW GUIDE	67

List of figures

Figure 1	Research Problem.....	4
Figure 2	Thesis Structure.....	6
Figure 3	Literature Categories	9
Figure 4	Framework for Systematic Literature Review	21
Figure 5	Final Search String.....	23
Figure 6	Master Thesis Data Collection Process.....	31
Figure 7	Risk Matrix.....	39
Figure 8	Model of The Norwegian National Security Authorities Basic Principles for ICT Security	41
Figure 9	Norwegian HelseCERT.....	44
Figure 10	Risk Assessments for Similar Equipment	48
Figure 11	Conceptual Model for Risk Assessment for Similar Equipment.....	49
Figure 12	Risk Management.....	50
Figure 13	Organizational-level Information Security Governance	52
Figure 14	Current State of Interoperability in the Norwegian Healthcare ..	53
Figure 15	Level-3 or Semantic Interoperability	54
Figure 16	Conceptual Framework for Ensuring Compliance.....	56

List of tables

Table 1	Summary of Security Challenges.....	3
Table 2	Research Question and Motivation	5
Table 3	List of Publications for the Literature Review	7
Table 4	Description of the five V's in Big Data	10
Table 5	Personal Health Information Governance in Healthcare	13
Table 6	Levels of Interoperability.....	15
Table 7	Core Features of PDR for this Master Thesis	20
Table 8	Search Strings.....	22
Table 9	Practical Screen	23
Table 10	Quality Appraisal Criteria	24
Table 11	Qualitative Data	25
Table 12	List of Informants.....	28
Table 13	Interview Analysis Process	29
Table 14	List of Publications for the Document Analysis	30

LIST OF ABBREVIATION

AWS	Amazon Web Services
BDA	Big Data Analytics
CERT	Computer Emergency Response Team
C.I.A	Confidentiality, Integrity, and Availability
EHR	Electronic Health Record
EIF	European Interoperability Framework
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
ICT	Information and Communication Technology
IoT	Internet of Things
IS	Information Systems
ISO	International Organization for Standardization
ISMS	Information Security Management System
IT	Information Technology
NIST	National Institute of Standards and Technology
NSM	Norwegian National Security Authority
OAG	The Office of the Auditor General
PDR	Phenomenon-Driven Research
PHI	Personal Health Information
PII	Personally Identifiable Information
SLR	Systematic Literature Review
TFA	Two Factor Authentication

1 INTRODUCTION

1.1 Research Background

Many institutions, as well as private individuals, are concerned about cybersecurity in 2022. Healthcare organizations are some of the entities we have the most faith in, and who have access to some of our most personal information, such as our name, date and place of birth, medical data, social security numbers, and so on. Even though health organizations possess our most sensitive and important information, they have become easy targets for hackers due to a variety of faults (poor budget, lack of IT organization, extensive use of legacy systems, and so on) (Le Bris & El Asri, 2016, p. 1).

The health sector and the health services has been undergoing an extensive change process during the last decades. The use of new medical technology, genetic mapping, artificial intelligence, and big data contribute to new opportunities for more effective treatment methods, as well as increased health research in many areas (Befring & Sand, 2020). Big data holds the promise of supporting a wide range of unprecedented opportunities and use cases, including the following key examples: For disorders impacting several organ systems, clinical decision support, health insurance, disease surveillance, population health management, adverse event monitoring, and treatment optimization (Abouelmehdi et al., 2017, p. 74).

In recent years, there has been a significant change in the sharing of personal health information (PHI) in the Norwegian healthcare industry. In August 2020, the construction of “Helseanalyseplattformen” (Health Analysis Platform) started, where data from all the different registers related to the Norwegian healthcare sector will be available in secure analysis platforms. As the Health Analysis Platform is developed, it will facilitate more advanced analyzes of Norwegian health data and lay the foundation for new types of medical and health research (The Directorate of eHealth, 2021). This solution will be the most efficient system for processing health data, and the one that will to the greatest extent contribute to more health research, innovation, and business development. Thus, the Health Analysis Platform will ensure a high socio-economic profitability (Åm et al., 2021).

“Helseplattformen” (The Health platform), is another large-scale project within the Norwegian Healthcare sector. This platform is set to launch in spring 2022 and is developed to be a new collaboration solution for the entire health service in

Central Norway, for the benefit of the patient. It is a large, joint project owned by Helse Midt-Norge and Trondheim municipality. In addition to a solution for patient records and patient administration, the Health Platform is a professional system that will support health personnel in patient treatment. With new and modern tools for sharing PHI, they can spend less time on paperwork and duplication in several different systems, which is a big-time thief today. All data is gathered in one place, and the inhabitants also get better access and overview through the patient portal (Helseplattformen, 2022).

To support the delivery of efficient and proper patient care, in addition to sharing data used for research purposes such as “Helseanalyseplattformen” and “Helseplattformen”, the Norwegian healthcare institutions and its stakeholders store, retain, and transfer massive volumes of data. Nonetheless, protecting these personal health records and personal identifiable information has shown to be a difficult task for decades worldwide. In fact, the healthcare sector and its Information and Communication Technology (ICT) platforms remains one of the most vulnerable to publicly revealed data breaches. Most of the threat actors normally utilize data mining methods and procedures to uncover sensitive information to make it public, resulting in a potential data breach (Abouelmehdi et al., 2017, p. 75). Seh et al. (2020, p. 4-5) states that from during the last 15 years, over 60 percent all data breaches are related to the healthcare sector. There has been a considerable rise in breaches in the healthcare sector in recent years, with over 75 percent of breaches occurring in the past five years (Seh et al., 2020).

As there has been a significant increase of data breaches within the healthcare sector worldwide, we have also seen the trend occur within the Norwegian healthcare sector. The Office of the Auditor General (OAG) is the Norwegian parliament’s auditing agency (the Storting). This organization is unusual in that it is the only institution capable of providing the parliament with a full and impartial government audit (The Office of the Auditor General, 2022). In 2021 OAG published the article “*Examination of the health-institutions’ prevention of attacks on their ICT systems*” where they describe how they simulated cyber-attacks against the Norwegian healthcare institutions resulting in a high degree of control over the ICT infrastructure in three of the four health regions. They also managed to gain access to large amounts of sensitive health information in all regions (The Office of the Auditor General, 2021, p. 5). The Directorate of eHealth in Norway published a report which focused on the overall risk and vulnerability assessment for ICT in the health and care sector. According to the research, various cases of flaws in the security culture of health institutions have been discovered, including a lack of knowledge and risk understanding of information security among management and personnel (The Directorate of eHealth, 2019, p. 26). This can be seen in light of the OAG report, which resulted in ethical security hackers gaining access to different sensitive data among the different health regions in Norway.

During the simulation the ethical hacking team got a huge amount of information about these different aspects of the healthcare systems in Norway:

- All accounts (user accounts, administrator accounts and service accounts).
- All access rights granted to accounts, including accounts granted to many rights.
- Systems that lack security updates and therefore have known technical vulnerabilities.
- Systems that are vulnerable due to errors in installation and operation.
- Older systems with fewer and weaker safety mechanisms.
- Several critical databases, for example electronic patient records.
- Multiple critical servers, such as file servers with sensitive information.
- Control panel for technical equipment (medical and building technology).

(The Office of the Audit General, 2021, p. 23).

The audit reveals that the health regions have taken several steps to improve information security. In recent years, health regions have attempted to upgrade or build regional information security management systems. Information security is the focus of several resources, and the professional settings of regional ICT providers have been enhanced. Major regional information security enhancement projects have also been developed, as well as new regional EHR for interaction. The table below illustrates how OAG sums up the most central challenges the Norwegian healthcare sector are facing.

Table 1 Summary of Security Challenges

Summary of security challenges in the Norwegian Healthcare Sector	
Main Challenge	Description of Challenge
Complexity and scope of equipment, systems, and software.	This allows for the correction of known flaws to take place over time. The health regions also consider that in some circumstances, outdated and ineffective technical solutions obstruct proper security.
Lack of cleanup.	The health regions prioritize the introduction of new solutions that will increase security, without phasing out the old, unsafe solutions. It is not systematically cleaned in old solutions and sensitive health and personal information. This can be exploited by attackers.
Unclear division of responsibilities and tasks.	There are ambiguities between the ICT providers and the health-institutions about who will implement information security measures, and in some cases disagreement about division of tasks.
Individual employees' security focus.	Both employees at the health-institutions and ICT personnel at the regional ICT providers

	<p>have a behavior that contributes to weakening ICT security.</p> <p>The information security training in the health-institutions is not adapted to the individual employee tasks.</p>
--	---

In light of the audit from the OAG, at the same time as major projects such as the Health Analysis Platform and the Health Platform in Central Norway will be launched, this master's thesis will go more in detail on how the health sector in Norway can avoid losing sensitive health data.

1.2 Research Problem

Big data is pervasive in the 21st century, affecting many parts of human existence, including biology and medicine. Furthermore, the migration from paper medical records to Electronic Health Records (EHR) systems has resulted in an exponential increase of data (Baro et al., 2015, p. 1). As a result, big data offers physicians, epidemiologists, and health policy specialists an unique opportunity to make data-driven decisions that will eventually improve patient care (Sessler, 2014, p. 104).

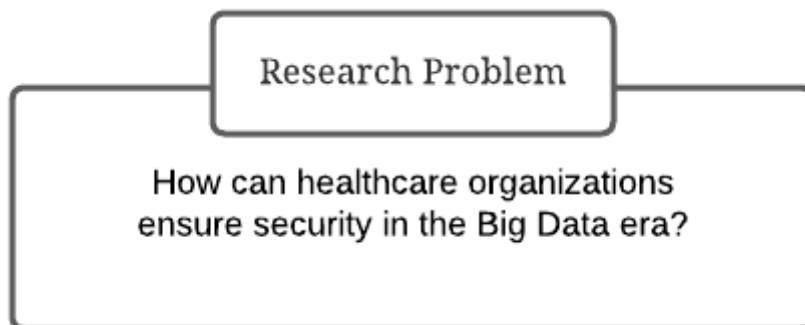


Figure 1 Research Problem

Based on the upcoming projects related to sharing health records between institutions in Norway, and the current state of cyber security within the Norwegian healthcare sector, this thesis will attempt to answer the research problem presented in figure 1 above.

1.3 Research Questions

In this thesis, the research questions include an assessment of the greatest challenges and concerns related to the security of Big Data in healthcare, as well as another question whose goal is to identify the primary security dimensions on which academics are focusing their efforts. Finally, we wanted to study about several strategies, methodologies, and models that have already been developed to address these issues. The research questions are used as a tool to better enable us to answer the main research problem stated in chapter 1.2. Table 2 below shows a definition of the research questions and the motivation behind them.

Table 2 Research Question and Motivation

Research Question	Motivation
RQ1: How can data governance framework mitigate security risks in the healthcare sector?	If the health data is available and maintains its integrity, this data can be a significant benefit for patient care and researching. However, due to the sensitivity of the data in health records, confidentiality must be considered when processing this data.
RQ2: How can Norwegian the Norwegian healthcare sector attain interoperability in integrated systems?	We aim to see if interoperability frameworks can be beneficial in order to maintain patient privacy when implementing systems for sharing and analyzing health data across health regions in Norway.

1.4 Thesis Structure

A literature review based on the research background and problems will be presented in Chapter 2. The research design and technique for the thesis are presented in Chapter 3. The findings from the empirical data, interviews, and document analysis will be presented in Chapter 4. The data will be discussed in Chapter 5 along with the theory from Chapter 2. The study's conclusion will be presented in Chapter 6. The study's limitations will be covered in Chapter 7, and future research will be suggested in Chapter 8.

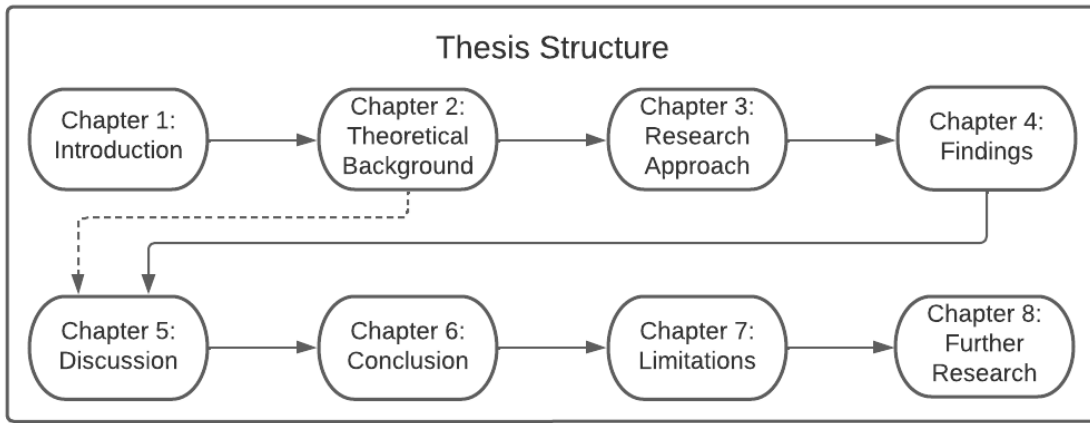


Figure 2 Thesis Structure

2 THEORETICAL BACKGROUND

This chapter provides an outline of relevant research for this master thesis. In addition, a literature review was undertaken to establish the current state of cyber security in the context of big data in the healthcare sector. Later in this chapter, a discussion of the literature review is presented.

2.1 Selected Literature

The articles in this section were chosen based on the criteria further explained in chapter 3.1. Articles and research that did not meet any of the requirements for inclusion criteria, were not included in the literature review. The following list can be used to match the writers mentioned in the table below with the titles of selected literature. A total of 22 articles were selected for this literature review.

Table 3 List of Publications for the Literature Review

List of Publications for the Literature Review			
#	Author	Title	Publication year
1	Hemingway et al.	Big data from electronic health records for early and late translational cardiovascular research: challenges and potential	2017
2	Lakshen et al.	Big data and quality: A literature review	2016
3	Gupta et al.	Big data with cognitive computing: A review for the future	2018
4	Cowie et al,	Electronic health records to facilitate clinical research	2017
5	Jaïdi et al.	Advanced techniques for deploying reliable and efficient access control: Application to E-healthcare	2016
6	Tipton et al.	Toward proper authentication methods in electronic medical record access compliant to HIPAA and CIA triangle	2016
7	Binjubeir et al.	Comprehensive survey on big data privacy protection	2019

8	Tse et al.	The challenges of big data governance in healthcare	2018
9	Jain et al.	Big data privacy: a technological perspective and review	2016
10	Patil & Seshadri	Big data security and privacy issues in healthcare	2014
11	Cavanillas et al.	New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe	2016
12	Al-Shomrani et al.	Policy enforcement for big data security	2017
13	Morabito, V.	Big data governance	2015
14	Alofaysan et al.	The significance of data governance in healthcare	2014
15	Trom & Cronje	Analysis of data governance implications on big data	2019
16	Winter & Davidson	Big data governance of personal health information and challenges to contextual integrity	2019
17	Juddoo et al.	Data governance in the health industry: Investigating data quality dimensions within a big data context	2018
18	Iroju et al.	Interoperability in healthcare: benefits, challenges and resolutions	2013
19	Salas-Vega et al.	Big data and health care: challenges and opportunities for coordinated policy development in the EU	2015
20	Dash et al.	Big data in healthcare: management, analysis and future prospects	2019
21	Ullah et al.	Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare	2017
22	Kouroubali & Katehakis	The new European interoperability framework as a facilitator of digital transformation for citizen empowerment	2019

2.2 Review Findings

The findings elicited from the selected literature shown in table 3, *list of publications for literature review*, is presented in this section. Since the literature covers many various aspects of big data in the healthcare sector, the findings were divided into multiple segments based on comparable topics. The following topics emerged as relevant and consistent throughout this literature review. The topic selected to help answer the research problem were: Electronic Health Records, Privacy in Healthcare, Big Data Governance & Interoperability Frameworks.

LITERATURE CATEGORIES

■ Electronic Health Records ■ Privacy in Healthcare
 ■ Big Data Governance ■ Interoperability Framework

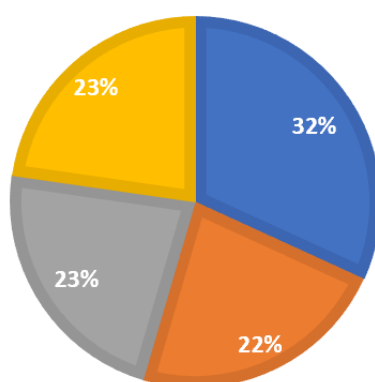


Figure 3 Literature Categories

Figure 3 provides an overview of how the literature is distributed into the different categories. Each category has at least five different academic sources. The further and more specified the topic got, the less relevant literature was to be found. This makes the research problem even more intriguing to investigate further.

2.2.1 *Electronic Health Records*

Variety, volume, velocity, and value are the four V's that define big data. There is also a fifth V, veracity, which is concerned with data quality (Hemingway et al., 2017, p. 1482). Both Lakshen et al. (2016, p. 2) and Gupta et al. (2018, p. 83) describes these five V's as follows:

Table 4 Description of the five V's in Big Data. (Lakshen et al., 2016; Gupta et al., 2018)

Description of the five V's in Big Data	
Variety	The richness of the data representation – text, graphics, video, audio, and so on – is measured.
Volume	The amount of data that is available to an organization; it does not have to possess all of it as long as it has access to it.
Velocity	A characteristic of big data in which it is vital to keep track of the pace at which data is generated while also being concerned with the speed at which that data is processed.
Value	The vast volume of data is worthless until it is converted to knowledge, according to this characteristic of big data.
Veracity	A feature of big data that deals with predicting the data's quality, uncertainty, and trustworthiness.

EHR are classified as Big Data in the context of this thesis's research problem because of its variety, number of patients and the volume of information on each patient. EHRs are gathered for a variety of purposes such as clinical care, billing, auditing, and quality monitoring (Hemingway et al., 2017, p. 1482). The purpose of EHRs is to improve patient care by integrating performance measurements into clinical practice and improving the identification and recruitment of eligible patients and healthcare professionals for clinical research (Cowie et al., 2017, p. 2). When processing such a large degree of sensitive information there is a vast need of having security measures in place to protect the data.

Given that security risks can have a significant influence on a patient's privacy, health, or life, security is a critical part of e-healthcare. The core of e-healthcare systems is confidential and sensitive data, which should be managed with extreme caution (Jaïdi et al., 2016, p. 1). Within the healthcare sector, Tipton et al. (2016, p. 5) argues that three goals exist as the benchmarks in the evaluation of information security: confidentiality, integrity, and availability – known as the C.I.A Triangle (Tipton et al., 2016)

There are several security measures that can be taken to protect EHR's. Tipton et al. (2016, p. 3-5) mentions different measures related to access control. The argument is that in order to maintain the confidentiality of the health records it is vital that only those who need and are authorized have access to the data. Measures

like two-factor-authentication (TFA), different standards of encryption, biometric authentication and anonymization are mentioned (Tipton et al., 2016; Binjubeir et al., 2019, p. 20071). The sensitivity of information in the EHR's makes it vital to understand some privacy issues related to them in order to understand how to govern them.

2.2.2 Privacy in Healthcare

Medical information stored digitally such as EHR is very sensitive and may infringe on an individual's privacy. If large data isn't properly governed, it can be easily abused by various threat actors. Before one can start using big data for the benefits of the analytical findings, attention must be paid to each stage of data modification to guarantee that the medical data is used correctly and efficiently (Tse et al., 2018, p. 1633).

Big data security and privacy are both huge challenges for both customers and service providers. In fact, in 2016, 80 percent of large-scale organizations had experienced severe security vulnerabilities related to big data. The majority of them are not in conventional formats, making analysis using today's technologies more challenging. According to reports, the rise of big data is increasing the risks to data security. One of the challenges with big data privacy is policy management, and how to enforce it with such enormous volumes of data without compromising speed (Al-Shomrani et al., 2017, p. 2).

Anonymization, for example, is a well-known technology used today for removing personal information about a patient's health. The problem with anonymization of data, is that anonymized data might be re-identified accidentally by combining huge data from several distinct data sources. As a result, current privacy-enhancing approaches must be assessed to see if they can fulfill all privacy criteria, even when working with large amounts of data such as EHR. If a technique such as data anonymization cannot ensure data privacy, it must be modified to meet the demand for privacy, or other methods and approaches must be devised (Cavanillas et al., 2016, p. 191).

Jain et al. (2016, p. 2) argues that the use of big data in healthcare raises security and patient privacy concerns dramatically. Patient data is initially housed in data centers with various levels of protection where, traditionally, security measures are ineffective when dealing with enormous data sets that are intrinsically heterogeneous (Jain et al., 2016). Patil & Seshadri (2014, p 763) further describes how the same problems with healthcare security and privacy is connected to data more than certifications and policies. In addition to patient data being stored in data centers with various security (Jain et al., 2016, p 4), most of the US healthcare data centers have the Health Insurance Portability and Accountability Act

(HIPAA) certification integrated, but however, this certification does not guarantee the security of patient records. The reason for this is that the certifications and principles is more concerned with guaranteeing security policies and procedures than with putting them in place. Furthermore, the flood of big data sets from many sources contributes to the storage, processing, and communication challenges (Patil & Seshadri, 2014, p 764).

With the rise in popularity of healthcare cloud solutions, the challenge of securing large distributed Software as a Service (SaaS) systems has grown, thanks to a variety of data sources and formats. As a result, prior to exposing data to analytics, big data governance is required (Jain et al., 2016, p. 19) (Patil & Seshdari, 2014, p. 764).

2.2.3 *Big Data Governance*

Morabito (2015, p. 85) defines data governance as:

“A system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods.” (Morabito, 2015)

According to Alofaysan (2014, p. 177), administrative, technological, and business modules are the three basic associated components of data governance in healthcare.

The Governors in a formal capacity are represented by the administration module, who are in charge of setting data governance scope and policies, resolving disputes, and allocating stewardship duties and tasks. The stewardships that are responsible for data standardizations and definitions, as well as compliance with data standards, business regulations, and procedures, are represented in the business module. Finally, the technical module represents IT employees who are in charge of data governance’s technical features, such as data integration rules and data modeling standards (Alofaysan et al., 2014, p.178-179). In healthcare, one of the key problems in generating value from big data has been establishing good governance. A successful data governance program encourages the creation of formal data standards and norms, as well as data supervision, so that decision makers have access to high-quality, consistent, and timely data to respond to the healthcare organization’s issues and opportunities (Trom & Cronje, 2019, p. 649). The advantages and difficulties of a successful data governance strategy are numerous and varied. People and procedures, rather than technologies, are at the heart of data governance concerns. Alofaysan et al. (2014, p. 179) argues that one of the most significant difficulties in healthcare data governance is a lack of business participation and executive level support. In reality, leadership

commitment is challenging because data governance is typically seen as a mystery issue rather than a program that provides business metrics in. The absence of people's knowledge of governance, which includes non-sustainable executive sponsorship as well as a lack of commercial reason, is the second most significant hurdle (Alofaysan et al., 2014, p.180). However, Winter & Davidson (2019, p. 38) discovered five analytic factors that describe distinct (PHI) governance forms and created a conceptual taxonomy of forms based on diverse arrangements of these dimensions. Table 5 below shows these different forms of PHI governance in healthcare.

Table 5 Personal Health Information Governance in Healthcare. (Winter & Davidson, 2019, p. 38).

Personal Health Information Governance in Healthcare				
Data domain	Stakeholders	Value/application	Governance goal	Governance form
Organizationally collected data , e.g., Individual's health history EMR (emergency medical responders) clinical encounter data Prescription/pharmacy data Lab data Imaging data	Direct , e.g., Individual Family members Health care provider 3rd party payers Employers Indirect , e.g., Government policy makers	Improvements via data analytics , e.g., Individual's health Organizational performance Health system's efficiency and effectiveness Evidence-based health services	Assuring and maintaining , e.g., Trust in governance Privacy Data security Regulatory compliance Facilitating , e.g., Data access Data analytics Innovations	Policies , e.g., Data privacy Regulations , e.g., EU's GDPR (General Data Protection Regulation) Organizational , e.g., Data access committee Neutral third-party data organization Technology , e.g., Algorithms Cyber-security tools Standards , e.g., Data harmonization via codes Interoperability protocols
Personally generated data , e.g., Activity data (e.g., diet, exercise) Clinical data (e.g., glucose level)	Heath researchers "The public" or "communities"	Community or population health Monetization of data	Protecting IP	
Digital trace data , e.g., Behavioral data from digital sources Online shopping Web searches	Health system , e.g., Health IT firms Pharmaceutical firms Medical equipment manufacturers			

There are various forms of PHI governance within the healthcare sector, which we can see in table 5. Organizational-level governance, in which a hospital or other clinical institution is the principal steward and consumer of PHI data created on its own ICT. Individual-level clinical data, along with operational systems such as clinical records and financial payments, are kept on file at the hospital. Through government requirements may compel data to be shared with accreditation agencies, researchers, or patients, this PHI data is "owned" by the institution (Winter & Davidson, 2019, p. 39).

Another form of PHI governance within healthcare are the individual-level. Individuals generate data in this setting by using consumer electronics like wearable activity trackers and glucose monitors, as well as entering data about their health-related activities into smartphone apps (data domains). Individuals and

the information technology IT firms who provide components or data aggregation services (stakeholders) share data governance rights and responsibilities, and data is typically stored on the IT vendor's cloud-based infrastructure as well as on the individual's mobile devices, such as a smartphone (governance form) (Winter and Davidson, 2019, p. 40).

Winter & Davidson (2019, p 40) points out that these different types of PHI governances within the healthcare sector often conflicts when data shifts from one context into another, particularly if informational practices, norms, and stakeholder values diverge from those in the original environment, such as when companies want to integrate consumer-generated health data with protected clinician health data onto their own technological platforms. Traditional health system stakeholders may face governance conflicts as well, such as between independent policy researchers focused on lowering health system costs through big data analytics (BDA) and clinical organizations whose economic and competitive interests may not be served by PHI-enhanced policy research (Winter & Davidson, 2019, p. 41).

One of the key problems in generating value from big data has been establishing good governance. A successful data governance program encourages the creation of formal data standards and norms, as well as data supervision, so that decision makers have access to high-quality, consistent, and timely data to respond to the healthcare organization's issues and opportunities (Trom & Cronje, 2019, p. 649). Trom & Cronje (2019, p. 649) further states that one of the “*major challenges that organizations face when trying to govern big data, is that big data is relatively beyond the organizational lines and mainly external*” (Trom & Cronje, 2019). As a result of issues such as conflicting data ownership and custody, regulating becomes a difficult undertaking. An organization can be controlled in a variety of ways. Strategies, goals, policies, plans, and standards are examples of governance systems. Juddoo et al. (2018, p. 3) argues that the main advantages for organizations that adopt a well-functioning data governance are:

- Increased value and revenue
- Managing cost and complexity
- Make certain of security, compliance, and privacy risk control.

(Juddoo et al., 2018).

Different governance processes are employed to create value and avoid risks, and governance has an influence on accomplishing an organization's strategic goals, as opposed to the managerial role, which focuses on attaining operational goals (Trom & Cronje, 2019, p. 650).

2.2.4 Interoperability Framework

The development of generally recognized standards can help to resolve interoperability concerns (Ullah et al., 2017, p. 4). Interoperability, in basic terms, is the capacity of multiple information and communications technology systems and software applications to interact, exchange data reliably, effectively, and consistently, and utilize that data (Iroju et al., 2013, p. 263). A key problem in the development of medical data systems is data interoperability is crucial for recording patient data, providing common interfaces, agreeing on similar data sets, and setting quality standards.. It implies the creation of data platforms in a global, comparable environment, which requires the use of common principles (Salas-Vega et al., 2015, p. 291). Patients may or may not receive treatment in more than one location. As a result, data exchange with other healthcare organizations would be critical. If the data is not interoperable during this exchange, data flow between various organizations may be severely limited. It is possible that this is due to technological and organizational obstacles (Dash et al., 2019, p. 20). Normally, there are 7 basic forms of interoperability (Iroju et al., 2013, p. 263) (see table 6).

Table 6 Levels of Interoperability. (Iroju et al., 2013, p. 263)

Levels of Interoperability	
Level 0 or no Interoperability	Stand-alone systems with no interoperability are typical examples of this.
Level 1 or Technical Interoperability	The adoption of a communication protocol for data transmission between systems is required at this level of interoperability.
Level 2 or Syntactic interoperability	This is the capacity of two or more systems to communicate data and services via the use of a common interoperability standard.
Level 3 or Semantic Interoperability	The capacity of two or more systems to automatically understand the information shared meaningfully and accurately in order to deliver valuable outcomes as defined by the systems' end users is referred to as semantic interoperability.
Pragmatic Interoperability	This degree of interoperability is reached when the interoperating systems are aware of each other's methodologies and procedures.
Dynamic Interoperability	When two or more systems are able to grasp the state changes that take place in the assumptions and constraints that they are making over time, and they are able to benefit of those adjustments, they are said to have achieved dynamic interoperability.

Conceptual Interoperability	If the assumptions and restrictions of the meaningful representation of reality are aligned, conceptual interoperability is achieved.
-----------------------------	---

However, in the context of healthcare and Big Data, Iroju et al. (2013, p. 264) argues there is no standard definition of interoperability. However, the National Alliance for Health Information Technology defines interoperability in the healthcare setting as the capability of multiple information technology technology systems to communicate, share data reliably, effectively, and consistently, and use that data.(Iroju et al., 2013). The interoperability of messages (information) exchanged between healthcare applications, the interoperability of Electronic Healthcare Records (EHR), the interoperability of patient identifiers, coding terms, clinical guidelines, and healthcare business processes are all examples of interoperability in healthcare. All of these interoperability characteristics, however, may be divided into two fundamental layers: syntactic interoperability and semantic interoperability (Iroju et al., 2013, p. 265).

Within the context of European public service delivery, the New European Interoperability Framework (EIF) defines interoperability as

“The ability of organizations to interact towards mutually beneficial goals, involving the sharing of information and knowledge between these organizations, through the business processes they support, by means of the exchange of data between their ICT systems” (Kouroubali & Katehakis, 2019, p. 3).

This interoperability framework provides fundamental interoperability rules for the delivery of European public services in the form of common principles, models, and recommendations. It promotes government agencies to create and offer services that are:

- Digital-by-default, with services and data delivered mostly through digital methods.
- By default, it is cross-border, and it is available to all EU citizens.
- By default, everything is open, allowing for reuse, collaboration, access, and transparency.
- Privacy-by-design and security-by-design infrastructure and building blocks that are compatible with legal data protection and privacy standards and duties.
- Interoperability-by-design as a framework for the design and delivery of European public services

Kouroubali & Katehakis (2019, p. 4) also states that the European Interoperability Framework will, in fact help organizations within the healthcare success when it comes to interoperability. In addition to EIF, healthcare organizations already have its own interoperability framework. This framework is called the Refined eHealth EIF (ReEIF), which explains six layers of interoperability, each with its own set of actors and activities. It can be used in

conjunction with the new EIF to help with healthcare interoperability (Kouroubali & Katehakis, 2019).

2.3 Discussion on the Literature Review

During this literature we gained a better knowledge of our problem area and research questions. Through the analysis, we have obtained an extensive comprehension of how Big Data within the healthcare sector can be a huge advantage for the society, both when it comes to patient treatment and medical research. As aforementioned, several high-priority projects in the Norwegian healthcare industry are on the verge of being implemented. Both “Helseanalyseplattformen” & “Helseplattformen” are projects that are being implemented, which include a high level of sensitive data/personal information such as EHR. Given the recent audits/reports on the security of the Norwegian healthcare industry, we felt it was vital to conduct additional research into how these projects could be performed without compromising the safety of patients and data.

The literature analysis revealed that the healthcare industry stores a significant amount of PHI, such as EHR, in various services, posing a significant danger to patients' security and privacy. The three main goals of evaluating the security in EHR are the C.I.A triangle. To be able to maintain as much C.I.A of the data as possible, there are several different security measures that need to be considered.

Salas-Vega et al. (2015, p. 290) states that there are five different key challenges that are frequently mentioned when it comes to the use of big data in healthcare: Confidentiality and data security, access to information, data reliability, interoperability, and management and governance (Vega et al., 2015). The research conducted on privacy within big data in healthcare clearly states that there are various techniques when it comes to maintain the privacy of the patients. Technical solutions such as data anonymization can be established and used with certifications and principles that are designed for health data. But some of these guidelines are more concerned about security policies and procedures than with putting them in place. Therefore, Tse et al. (2018, p. 1633), Jain et al. (2016, p. 19) & Patil & Seshadri (2014, p. 763) all argues that establishing a good data governance is crucial to maintain the privacy within EHR.

Big Data Governance is crucial for any stakeholder within the healthcare sector to maintain already established procedures, formal standards and norms, and technical solutions so that the so that decision makers have access to high-quality, consistent, and timely data to adapt to the challenges and opportunities faced by the healthcare organization (Trom & Cronje, 2019, p. 649). This is done to keep the data confidential (privacy for the patient), maintain the integrity of the data (to

be able to perform best possible patient care or research) and availability (most important in life-saving situations for the patients). Based on reports indicating how vulnerable the Norwegian healthcare sector is to cyber threats, we believe it is critical to focus on data governance, particularly in light of upcoming projects in which data will be shared on a large scale among various stakeholders, including hospitals, research institutions, patient applications, and general physicians (GP). As presented in table 5, the (*Data Domain*) is stored/shared between different (*Stakeholders*), and this data will be used to improve patient care using BDA and other methods (*Value/application*). To be able to perform analytics and patient care the best possible way, (*Governance goals*) should be established. Lastly, the (*Governance form*) should be well defined stating who is responsible for the different aspects of governance, such as policies, regulations, technological aspects, standards used and organizational structure.

According to the research done in this literature review on Big Data in the healthcare sector, Data Governance is crucial for any organization handling EHR, but when the data is transferred between different stakeholders, well established Data Governance will not be enough to completely secure the data. To keep the data confidential & maintain the integrity, the use of an Interoperability Framework is suggested as a method. In the Norwegian healthcare sector, there are several different health regions, where each one of them has their own Data Governance. To be able to transfer, share and use EHR in Norway (and to other countries in Europe or third-party countries), we find it important for these institutions to have a close to similar interoperable framework. A framework based on the EIF and ReEIF, which contains different layers that each stakeholder must follow to be able to store, transfer and analyze EHR will ensure that the patient privacy is considered, and the electronic platforms are more secure.

In summary, privacy, data governance and interoperability frameworks are what we find most important to consider while storing, transferring, and analyzing EHR. Based on the recent reports about how vulnerable the Norwegian Healthcare sector is, as well as how the healthcare sectors ICT platforms remains one of the most vulnerable to publicly revealed data breaches (Abouelmehdi et al., 2017, p. 75), we find it necessary to further investigate how the Norwegian healthcare sector works. This forms our foundation for further research to identify the processes when EHR is transferred between different stakeholders in Norway, and how integration of data governance and interoperability frameworks will strengthen the security of the patient's health records.

3 RESEARCH APPROACH

The term "research approaches" refers to a set of study plans and procedures that span everything from broad assumptions to specific data collection, analysis, and interpretation methods. The overall decision involves deciding which strategy to use to research the subject of this thesis. Research approaches are study plans and procedures that cover everything from general assumptions to detailed data collecting, analysis, and interpretation methodologies. The overall decision entails which method should be employed to investigate the subject in this report (Gregar, 1994).

Despite the importance of research in both commercial and academic operations, there is no agreement in the literature on how it should be represented. This might be due to the fact that different people perceive research differently. However, there appears to be agreement among the many different definitions that:

- Research is a process of enquiry and investigation.
- It is systematic and methodical; and
- Research increases knowledge.

(Amaratunga et al., 2002).

3.1 Phenomenon-driven Research

This master thesis is based on a problem-oriented research approach called Phenomenon-driven research (PDR), which focuses on capturing, documenting, and understanding an observable phenomenon of interest in order to assist knowledge generation and development (Schwarz & Stensaker, 2016, p.1). The table below is based on Schwarz & Stensaker (2016) table "Core features of a PDR paper", which represent the structure on this thesis.

Table 7 Core Features of PDR for this Master Thesis

Core features of PDR for this Master Thesis		
Core Features	Characteristics	Key Features
Aim and motivation. (Background)	Understanding the phenomenon	Investigating existing audits on the Norwegian health sector (Empirical story of the industry)
Audience and goal of research. (Research Problem and Research Questions)	The audience of this thesis is academics and practitioners. The goal is therefor to bring fresh perspectives and insights for academics and practitioners.	Framing the research problem and its following research questions in an empirical manner.
Role of theory. (Systematic Literature Review)	Using theory to place the study and phenomenon in context or to construct new theories to describe and explain phenomena, eclectically drawing on and integrating multiple theories is used.	The phenomenon and research topic are theoretically framed with literature collected in the literature review. Further on, a discussion based on the literature was done to summarize the different aspects.
Research methods. (Interviews and Findings).	Qualitative research method is chosen in this section.	Semi-structured interviews are performed to gain a better understanding of the phenomenon. Document analysis is used with analyzing the interviews to get a better understanding of the research problem.
Contribution to knowledge. (Discussion & Conclusion).	Developing a model or framework for describing and interpreting the phenomenon by mapping structures onto a new phenomenon or new constructs onto an existing phenomenon.	A framework based on the findings from SLR, background and interview findings are presented to better understand the phenomenon.

3.2 Literature Review

To get an understanding of the related research already conducted on the study area it was chosen to use a Systematic Literature Review (SLR) for this research. As a guidance on how to conduct such a review, this thesis uses the process suggested by Okoli & Schabram (2010) in the paper “A Guide to Conducting a Systematic Literature Review of Information Systems Research.” This framework is made for research within Information Systems and the method used for this review is therefore adaptable to the context of this study (Okoli & Schabram, 2010).

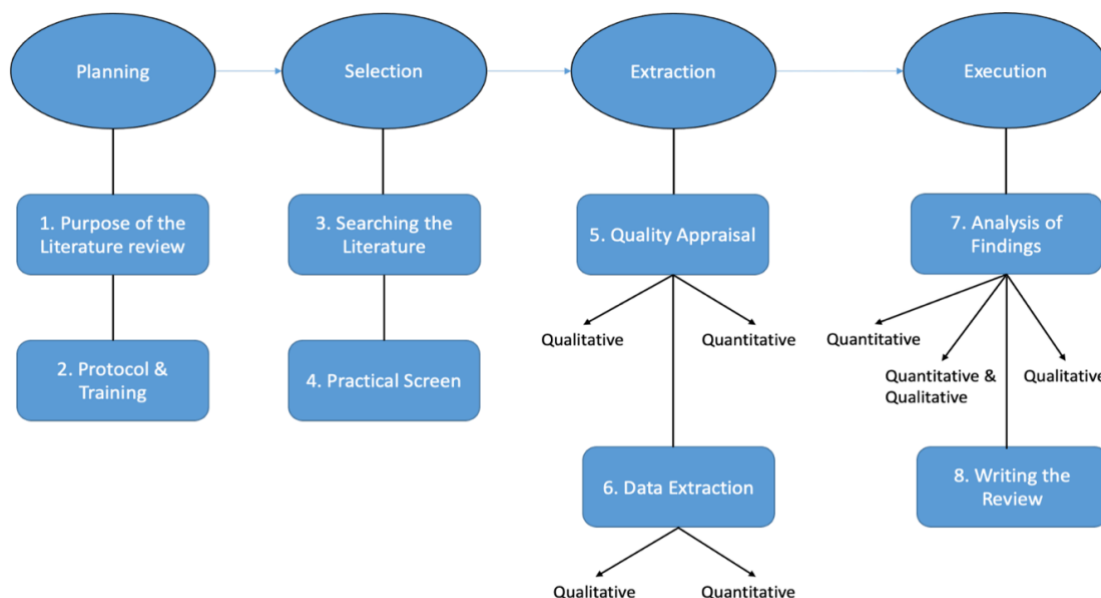


Figure 4 Framework for Systematic Literature Review

Figure 4 illustrates the framework used when conducting the literature review. The review has four different phases where each phase has two subcategories connected to them. Both qualitative and quantitative studies can be included if they meet the criteria for the literature and if they are of relevance to the study. Below is a summary of how each of these subcategories has been conducted to write a SLR for this thesis.

1. Purpose of the literature review

The purpose of the SLR is to get an understanding of the research that has already been conducted in relevance to the research problem presented in an analytical and systematic manner. In this SLR it is relevant to look at research related to how Big Data is used in relation to EHR and how data governance and interoperability frameworks can strengthen the patient privacy. The review gave valuable insight into how the healthcare sector use Big Data related to privacy and frameworks for how take advantage of information security in the sector. It also provided knowledge of the research area that can be used to outline the issues related to governance of data exchange across different organizations within the healthcare sector.

2. Protocol and Training

During the literature in this master thesis the protocol for choosing and writing the literature has been followed. The protocol is based on the many review criteria and specifies the particular stages and procedures to be followed. It was also critical

that both group members followed these criteria to guarantee that both were on the same page about what was being done. The protocol has therefore in this review been using the search strings through our chosen search engines, check if the literature is up to date with the practical screening criteria, and lastly evaluating the selected literature through the quality appraisal. Further down in this document each criteria is described as well as how we are following them.

3. Searching the Literature

During this process of writing this SLR, relevant literature has been found by searching through Google Scholar & Scopus. These search engines were used as they combine literature from numerous areas and provide a vast library of sources giving us sufficient results. During the search, all the relevant literature found was stored in a shared folder/document using a cloud service provider. The following strings were used while searching for relevant literature:

Table 8 Search Strings

SEARCH STRINGS
Big Data Security Healthcare
Big Data Security Interoperability Frameworks Healthcare
Big Data Security Interoperability Framework
Data Governance Healthcare
Big Data Governance Healthcare
Big Data Governance
Healthcare Privacy
Big Data Privacy
Privacy Governance Healthcare

The reasoning behind having these search strings is that big data, security, data governance and privacy are all connected to the data of which the research problem are exploring. The data in healthcare platforms explored in this thesis considers around big data. Since the information here is to be treated as sensitive information there is a strong need to be compliant with both privacy regulations and data security when working with data governance.

4. Practical Screen

After the search strings was defined there were several other criteria to be utilized in the review process. These criteria are more of a practical matter for the search process and are used to include and exclude literature for the search. The criteria are shown in table 9 below.

Table 9 Practical Screen

PRACTICAL SCREEN	
Publication Language	English
Date of Publication	2011 or newer
Duplication	No duplicate literature
Citations	Literature with no citations will be excluded unless they are from 2021 or newer
Citation Ranking	More referenced articles on similar subjects will be prioritized
Publication	Only final publications will be included
Subject Area	Computer Science & Social Sciences

Using the search strings combined with the criteria in the practical screen provides the following final search string for the literature:

Final Search String
(TITLE-ABS-KEY("Big Data Security Healthcare" OR "Big Data Security Interoperability Frameworks Healthcare" OR "Big Data Security Interoperability Framework" OR "Data Governance Healthcare" OR "Big Data Governance Healthcare" OR "Big Data Governance" OR "Healthcare Privacy" OR "Big Data Privacy" OR "Privacy Governance Healthcare")) AND (LIMIT-TO(PUBYEAR, 2022) OR LIMIT-TO(PUBYEAR, 2021) OR LIMIT-TO(PUBYEAR, 2020) OR LIMIT-TO(PUBYEAR, 2019) OR LIMIT-TO(PUBYEAR, 2018) OR LIMIT-TO(PUBYEAR, 2017) OR LIMIT-TO(PUBYEAR, 2016) OR LIMIT-TO(PUBYEAR, 2015) OR LIMIT-TO(PUBYEAR, 2014) OR LIMIT-TO(PUBYEAR, 2013) OR LIMIT-TO(PUBYEAR, 2012) OR LIMIT-TO(PUBYEAR, 2011)) AND (LIMIT-TO(SUBJAREA, "COMP") OR LIMIT-TO(SUBJAREA, "SOC")) AND (LIMIT-TO(LANGUAGE, "English")) AND (LIMIT-TO(PUBSTAGE, "final"))

Figure 5 Final Search String

The result in the database Scopus provided 222 results that matched the criteria. In addition to using Scopus as a search engine for this literature review, we also used Google Scholar. It is not possible to combine all the strings in one search in Google Scholar such as we did in Scopus, but the same strings and criteria has been used while searching in Scholar as well. In order to select the most relevant results for this thesis a standard for quality appraisal was established.

5. Quality Appraisal

Finally, criteria were established for evaluating the quality of selected literature. If the literature did not meet these quality assurances seen in table 10 below, they would not be included in the review.

Table 10 Quality Appraisal Criteria

QUALITY APPRAISAL CRITERIA	
Question	Answer needed for inclusion
Is the literature in accordance with the selection criteria?	YES
Is the literature relevant to the research problem?	YES
Is the study reliable in form of methodology used?	YES
Is there a clear bias in the literature?	NO

Some of the literature that met the initial criteria regarding search string and practical screen did not pass the quality appraisal. For instance, some research article focused more on the technological aspects of blockchain and internet of things (IoT) within the healthcare sector, which was not deemed relevant for this research. There were also some studies that did not define how the research was conducted and thereby excluded from the literature review.

6. Data Extraction and Synthesis

The data extraction worked in the way that sections from the literature that was deemed relevant for this study was extracted and put into and organized in an external document. No raw data of the extraction itself is included in this research but the summary and synthesized version of the findings with references to its sources is described earlier in chapter 2.1.

3.3 Qualitative Research Approach

The term "qualitative research method" is a broad term that covers a wide range of approaches and ideas, making it difficult to define. In general, qualitative research is a method of exploring people's experiences through a range of approaches such as in-depth interviews, focus group discussions, observation, content analysis, visual methods, and life histories and biographies (Hennink et al., 2020, p. 10). When compared to traditional quantitative data collection, qualitative methodologies generate a massive amount of data, and making sense of pages and pages of interviews and field notes can be difficult. Data organization and interpretation may appear to be a daunting task (Patton, 2014). This method is still used by some qualitative researchers to fulfill this challenging task. However, a number of well-known qualitative theorists have advocated for the use of qualitative data analysis software tools to assist researchers in handling data throughout the research process (Denardo, 2002. p. 2)

Table 11 Qualitative Data (Hennink et al., 2020, p. 16)

Qualitative Data	
Objective	Gain a contextualized understanding of behaviors, beliefs, motivation.
Purpose	To understand why? How? What is the process? What is the influence or context?
Data type	Textual data
Study population	Small number of participants; selected purposely (nonprobability sampling)
Data collection method	In-depth interviews, observation, group discussions.
Analysis	Interpretive analysis
Outcome	To develop an initial understanding, to identify and explain behavior, beliefs, or action.

The goal of this thesis is to provide answers to the research questions and thereby be able to give a comprehensive analysis and discussion to find a solution to the research problem. To be able to achieve this there is a need for a research approach to the study that will provide the best possible outcome. Over the course of this study there is a need to get in depth knowledge of information security within the Norwegian healthcare sector. There are a lot of information that is publicized but that information is not comprehensive enough to answer the questions related to information security in the healthcare sector. The be able to answer these questions there is a need to know what has been done regarding security measures and evaluations related to the Health Analysis Platform, how the organizations within the healthcare sector are adapting to an ever-changing threat landscape, and where the security provides limitations to the applications' functionality. To be able to answer these questions, a qualitative approach is deemed most suitable and will be the research method for data collection used for this study.

3.4 Research Interviews

Qualitative research interviews can be separated into three different categories:

- Unstructured interviews
- Semi-structured interviews
- Structured interviews

The different types of interviews have different purposes and thereby different strengths and weaknesses depending on the setting of which they are conducted.

The unstructured interview is an interview in which the questions asked from the interviewees to the objects are not planned. The results of the interviews are

that they are spontaneous and if there is more than one person that are being interviewed the questions may vary from each interview. This will make it more difficult to compare results from different interviews. On the other hand, the interviews have less restraints and give results in form of experiences from the interview object (Chauhan, 2019, p. 1-3).

With semi-structured interviews the goal is to create a situation of which there is a free conversation around some specific themes brought up by the researcher. The interview objects will usually reflect around their own experiences and/or opinions around the theme. The questions are open which is ideal in order to make the interview objects go in-depth if they have a lot to talk about a certain theme. As with the unstructured interviews, some questions can be spontaneous if the interviewee finds it interesting to know more about a specific subject that is brought up in the conversation. The semi-structured interview is a combination between unstructured and structured interviews (Tjora, 2021, p. 127-128).

The structured interviews are interviews of which all the questions and the order of the questions are planned and asked in the same order for every person interviewed. The goal is to get an answer on that specific question and the questions are not open as in the semi-structured and unstructured interviews. This interview form limits the interviewee from having any form of improvisation or deviate from the questions in the interviews. This form of interview can be good in a setting where one will compare different answers on the exact same questions (Myers & Newman, 2007, p. 4).

3.5 Document Analysis

Document analysis is a type of qualitative research in which documentary material is analyzed and particular research questions are answered using a systematic approach. Document analysis, like other qualitative research methodologies, necessitates frequent inspection, evaluation, and interpretation of data in order to obtain meaning and empirical knowledge of the phenomenon that is being studied (Gross, 2018, p. 2).

In this research, document analysis is used as a supplementary research method to either fulfill the findings in the interview or find other perspectives compared to what the informants have given. It will also be used to go more in depth into certain subjects that is deemed interesting to investigate further from the interviews. The document analysis will be used in chapter 4 alongside information provided by the informants. These findings will be further discussed with the findings from the literature review in chapter 5.

3.6 Data Collection Process

The most suitable data gathering method for this study was determined to be a semi-structured interview approach combined with document analysis. The reason for choosing this format is that there is a need to clarify some existing issues, as well as a limitation on the researchers' in-depth knowledge of the research problem. This provides the necessity of asking follow-up questions into interesting areas that the informants provide in their answers. The semi-structured interview allows such a process with a general structure with predefined questions as well as space for following leads during the conversation between the interviewer and interviewee (Magaldi & Berler, 2020, p. 4825).

The interview guide was created with the idea of performing the interview as a more open conversation. The intention was to have around five pre-defined questions for the interviewee to discuss and some potential follow-up questions. This way it was possible to explore the area without setting too much limitation on the interviewee. As the answers unfolded during the interview the researchers could explore the fields of which was found interesting and thereby gain a broader knowledge of the study area. The interview guide can be found in appendix A. The selection of interview subjects is further described in the next sub-chapter 3.6.1 Selection of Research Subjects.

3.6.1 Selection of Interview Subjects

After the format of the interview was decided there was a need for a process to select candidates for the interviews. This research has a focus on governance and interoperability from a security perspective within the Norwegian healthcare sector. With that in mind it was clear that the subjects for the interview should either be working within, have previous experience or in-depth knowledge of the IT infrastructure in the healthcare sector. The candidates should also have some knowledge from a managerial perspective as the research problem has an angle on strategic leadership within information security. When the criteria were established, the next step was to contact potential interview candidates. Some of the candidates were contacted through Eirik Thormodsrud in Sopra Steria which helped us a lot in the beginning of the interview process. To reach out to new candidates, the candidates already interviewed provided contact information to other persons of interest for the research. In addition to that, e-mails and LinkedIn messages were sent to organizations and individuals interesting for the research. After the research was conducted, a total of seven informants participated in the research (see table 12).

Table 12 List of Informants

List of Informants		
ID	Position	Sector
1	External Consultant	Private
2	External Consultant	Private
3	External Consultant	Private
4	CISO	Public
5	Security Advisor	Public
6	CISO	Public
7	Information Security Manager	Public

3.6.2 Analyzing Interviews

This research looked at a range of data and structurally examined it. The analysis was based on interviews, where we took notes during the discussion and transcribed the audio for further analysis. Furthermore, we used an inductive approach to analyze the results data, and we chose to apply the form of thematic analysis explicitly (Thomas, 2006, p. 237). The following are some of the motivations for the creation of the general inductive analysis method based on Thomas (2006):

- “1. To condense extensive and varied raw text data into a brief, summary format.*
- 2. To establish clear links between the research objectives and the summary findings derived from the raw data and to ensure that these links are both transparent (able to be demonstrated to others) and defensible (justifiable given the objectives of the research).*
- 3. To develop a model or theory about the underlying structure of experiences or processes that are evident in the text data.”*

(Thomas, 2006, p. 238).

In combination of thematic analysis, we will use an inductive approach to analyze the interviews. Thomas (2006, p. 238) described deductive approach as: *“An inductive approach is a systematic procedure for analyzing qualitative data in which the analysis is likely to be guided by specific evaluation objectives”* (Thomas, 2006)

A thematic analysis combined with an inductive method will seek to address our research questions by presenting a model of how a data governance and interoperability framework can increase patient privacy in EHR. During this

process we have used the software NVivo to code the data. A further explanation of the analysis is presented in the table below.

Table 13 Interview Analysis Process

Interview Analysis Process	
Initial reading	First, we will start by cleaning up the raw data files and then reading them to see if there are any patterns or general ideas.
Coding Process	We were able to code each segment of the interviews with the responses from all of the interview objects through using the interview transcription tool NVivo. Using this strategy, we were able to compare each informant's similarities and differences to the appropriate interview question.
Coding into themes	After that, we began to code the transcriptions into themes in order to avoid having to use phrases that were too wide or comprehensive. This made it easier to find specific phrases or terms that would assist us comprehend what the informants had stated.
Translation	All informants were interviewed in Norwegian, and the information acquired from the interviews was subsequently translated into English.

3.6.3 Document Selection

The documents used in this part of the research is to either check against what the informants in the interviews told or to supplement the answers they provided. The documents are not just academic research publications, which is the case in the literature review. Some are directly linked to different standards and information from different institutions. One example of this is ISO 27001 which is gathered through "Standard Norge" which is the Norwegian member of the European Committee for Standardization and ISO. The publications listed are both in English and Norwegian.

Table 14 List of Publications for the Document Analysis

List of Publications for the Document Analysis			
#	Author	Title	Publication year
1	The Norwegian Government	New Personal Data Act	2019
2	Gisle	The Privacy Ordinance	2018
4	The Office of the Auditor General	Riksrevisjonens undersøkelse av helseforetakenes forebygging av angrep mot sine IKT-Systemer	2021
5	South	Scaling a governance, risk, and compliance program for the cloud, emerging technologies, and innovation	2018
6	Microsoft	Microsoft 365 Risk Management program	2021
7	University of Bergen	Risk Matrix	2021
8	ISO	About us	2022
9	Standard Norge	NS-EN ISO / IEC 27001 Management systems for information security	2022
10	IT Governance	ISO 27001 vs. ISO 27002: What's the difference?	2021
11	Aspøy	Nasjonal Sikkerhetsmyndighet	2022
12	Norwegian National Security Authority	NSM's grunnprinsipper for IKT-sikkerhet	2020
13	Microsoft	Windows Security	2022
14	Norwegian National Security Authority	Hendeshåndtering	2020
15	Norwegian Health Network	HelseCERT	2022

3.7 Master Thesis Data Collection Process

The research approach is described in the chapter above. The figure below (see figure 6) is an illustration of the data collection process we have used in this master thesis. The first step of this thesis is the background, where we did research on the current situation on the Norwegian Healthcare Sector, which included audits and

information about upcoming projects related to big data. Further on we did a systematic literature review on the existing theory on the research background. We then chose to conduct a qualitative research approach containing semi-structured interviews where we established an interview guide and selected informants. The data collected in the interviews was then coded through NVivo transcription tool. An analysis of the interviews with a document analysis to validate the data was done before we could make a discussion on the research done.

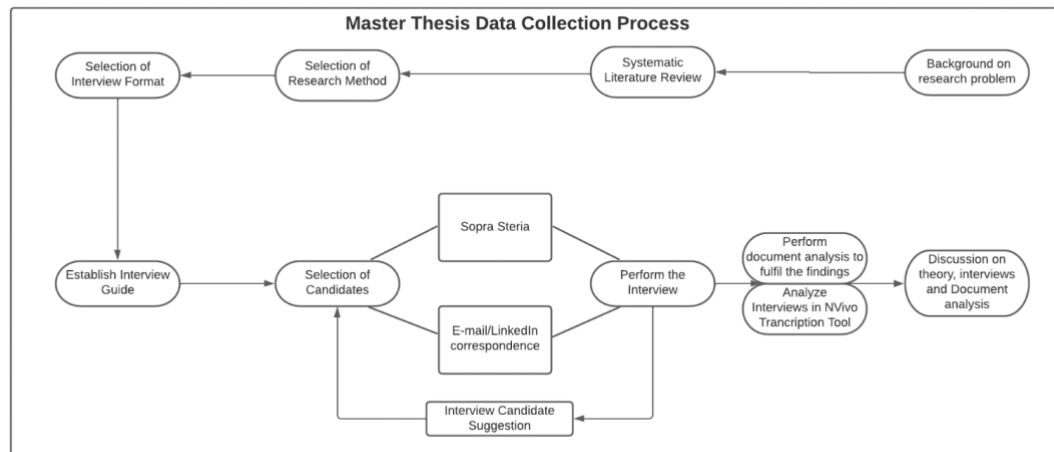


Figure 6 Master Thesis Data Collection Process

3.8 Validity & Limitation

In qualitative research, validity refers to the "suitability" of the instruments, methods, and data.

To ensure validity of the study several criteria must be met:

- The research question must be valid for the desired outcome
- The methodology must be appropriate for answering the research question
- The design must be appropriate for the methodology
- The sampling and data analysis must be appropriate
- The results and conclusions must be appropriate for the sample and context.

(Leung, 2015, p. 325)

As for this research, the validity of the study has been documented in previous chapters. Following the documentation for the choices made related to the research problem and research methodology, these criteria for validation of the research is met to the best of the researchers' knowledge. It is critical to point out that the study has some limitations. The interview questions are designed to eliminate bias to the best of the researchers' ability. However, it is possible that the questions, as

well as the additional follow-up questions, were biased as the dialogue progressed. The responses made by the participants might potentially be biased. There is a chance that the informants have forgotten important details that may be useful to the study, and that they have given a misleading depiction of their responses in some situations. The questions are open and can be interpreted differently by each informant to avoid bias in both the questions and the replies. The interviews are limited to seven informants as there in last couple of interviews was a distinct element of saturation in the answers provided.

3.9 Ethical Considerations

The Norwegian Center for Research Data (NSD) authorized the data collecting procedure and storage of the data. It was determined that neither the informants' names nor the organizations for which they worked should be made public. Each informant's position, as shown in table 12, was the sole piece of information directly related to them. The data would be kept on the University of Agder's cloud service, and only the researchers would have access to it. Throughout the study procedure, the informants had the option of having their information and responses erased at any point. As a result, the informants' identity has been preserved to the degree practicable in the context of writing a master's thesis.

4 FINDINGS

The findings from the interviews and document analysis will be covered in this chapter. The interviews have been coded in Nvivo and segmented into different categories. Each will provide valuable information to help provide a solution to the research problem stated in chapter 1.2. To validate the data, quotes from interviews and document analysis are provided.

4.1 Big Data in Healthcare

We would not have debated Big Data in the health sector if it weren't for all the opportunities it offers. Every informant we interviewed mentioned several different opportunities when it comes to Big Data in healthcare. Big Data opens up for greater access to important information, and for researcher to link data from the many registers we have in Norway. Today, obtaining health data is too complicated and time consuming. ID1, ID4, ID5 & ID6 all mentioned research projects and its possibilities while talking about opportunities with Big Data.

“The possibilities with big data in healthcare are that it can drastically shorten the time it takes for a research project” (ID5).

“So at least Big Data provides a lot. It can also support in such a treatment situation both in a curve system and a patient record system during decisions. If you have data related to a patient or a group, you can draw conclusions about the course of the disease, what works and what does not. It has been said at least (without me being a doctor), that big data provides better decision support in practice” (ID6).

The opportunities with Big Data in the healthcare sector are almost limitless. Better research and patient care, earlier detection of diseases like cancer and easier distribution of information are all part of it. However, processing a large amount of information provides several challenges related to confidentiality, integrity, and availability. One of these challenges are patient privacy.

4.1.1 Privacy

There are several barriers to exploiting the great opportunities big data can provide researchers and physicians, such as privacy, cloud storage, access control and various threat actors. One of the biggest challenges with Big Data in healthcare is privacy. All the informants mentioned privacy in different contexts during the interview, and informant ID6 described privacy and information security, and how they often connect in healthcare as follows:

“One thing is information security, but then you also have privacy that comes in strong in connection with this. And it is often connected. It is information security that ensures good measures on privacy” (ID6).

Informant ID2 further stated that:

“The GDPR says that everything that is health information is a special-category of information and they are all sort of the same category” (ID2).

General Data Protection Regulation (GDPR) is a regulation that was incorporated into the EEA Agreement by decision of the EEA Committee on 6 July 2018. Later in July 2018 the regulation came into force in Norway (The Norwegian Government, 2019). At the same time, the older Personal Data Act of 2000 was repealed. The purpose of the Regulation, as set out in Article 1 in GDPR, is to ensure the protection of the fundamental rights and freedoms of natural persons (individuals), and in particular their right to the protection of personal data. It also follows from Article 1 that the greatest possible free exchange of personal data between countries is desired. Privacy should therefore not lead to restrictions in the exchange of such information (Gisle, 2018).

The different health organizations in the different health regions in Norway uses different security measures to be able to use Big Data in compliance with GDPR. Four out of seven informants told us that pseudonymization was one of the most important measures they took to protect the patient privacy. ID4 stated that personal identifiable information such as name and social security numbers are stored separately and encrypted from other information. But the informant also stated that it can be difficult to ensure the privacy for a patient due to the fact that it can be easy to combine separated data together when there are small data sets, such as records from small municipalities. Informant ID1 mention the same problem with anonymizing data. The informant elaborated around how there are many ways of identifying a person without any name or social security number.

“We have had a lot of cases where you can find out that a person has been ill in a small village, then everyone knows this anyway, because the village is so small. You get so much data, you get new challenges around security and identification security, you can find connections without the patient's name. So there is a thing around

it with large amounts of data, one can get identification problems even if one has removed the identity" (ID1).

In addition, informant ID5 told us about how they anonymize the data, and why they do it. The informant walked us through how they were implementing systems with sensitive data, and how they used privacy principles to be in compliance with GDPR. The informant further elaborated about how researchers and doctors/nurses can gain access to the sensitive data. In many cases, researchers would need the data to perform their research, and to gain access to these data, an application must be sent to the authorities in Norway. The researchers then get access to anonymized data. According to the informant, they can acquire access to indirect identifying information in most research initiatives containing sensitive material. According to the informant, the researchers can acquire access to indirect identifying information in most research initiatives containing sensitive material. Lastly, informant ID6 also touched upon research project and anonymizing of data, where he stated that:

"You make moves in relation to pseudonymization (common in research) is mostly done. Then you have a number of security mechanisms. Both personnel security and physical security and access control, encryption" (ID6).

While most of the informants mention research projects, access control & anonymization, ID7 is more concerned about the type of sensitivity. It is mentioned that in psychiatry, for example, there are extremely sensitive information processed, which is by law deemed to be separated for different employees at the same workplace. ID7 further explains how it is difficult to find a balance between patient privacy, service needs for employees and the urgency that may occur in certain situations.

To summarize, the data collected through the interviews stated that privacy plays a big role in healthcare and big data. One of the key solutions to the patient privacy is anonymization. Anonymization can be used to strengthen the patient's privacy; however, it might not always be the solution. There is need to prioritize the data based on the type of health data, the use area, which can be either for research or patient care, and lastly the type of urgency of patient care in certain situations. Based on this, anonymization might not be enough in certain situations, due to the size of the population and data sets in the different health regions and the type of use area.

4.1.2 Other Challenges

It is clear from the research that many of the challenges of processing sensitive health information has its origin with laws and regulations regarding privacy. However, that does not mean that the privacy itself is the only challenge, but rather

that the solutions to these issues are not good enough to handle the privacy laws and regulations. As ID7 states:

“It is the entire privacy regulations that are risk-based. It is very leading. In that sense, I do not think it is the regulations that are in themselves a challenge. What is challenging are the large American suppliers and US authorities who have slightly greater rights to information than Europe has” (ID7).

This statement correlates directly to the fact that the major cloud delivery services are located in other regions outside the EU/EEA. ID5 further agrees with ID7 that the lack of European competitors in the cloud environment purposes a challenge. Had there been European cloud services that provided the functionality and security mechanisms as other big cloud providers, some of the privacy challenges facing the health sector would be close to eliminated.

There are also other challenges related to Big Data in healthcare. The challenges already elaborated for have to do more with strategy and compliance with the GDPR for healthcare systems. However, as ID5 points out, there might be security risks related to the open-source code structure itself.

“I would say not only when we develop, but the whole life cycle in fact, then we have to have possible compromised code (...) Many people can check the code when it comes in, but the open-source project itself may be compromised, or someone may replace components. In addition, there are known published vulnerabilities” (ID5)

In addition to concerns related to cloud service providers in different regions outside EU/EEU, there is also challenges internally in Norway. One of the big challenges in the current systems is that the data is widely spread over a series of different journals and systems in different regions across the country. Both ID4 and ID7 thinks the current system structure is at the expense of the availability and integrity of the data. They are both more unsure about the confidentiality but are clear that this is less important in the context of healthcare. As ID4 states:

“I think if you look at patient safety then integrity and availability are much more important than confidentiality. If you are in hospital and need surgery, the most important thing is that the medical records are correct and available. This is more important than unauthorized individuals seeing it [the journal]” (ID4).

The data collected shows that the main goal for data security is to improve patients' safety. This is the number one priority for the healthcare sector. This means that the issues related to security are managed in a way that focuses first and foremost on patient safety and then concentrates on other issues. However, this may vary from case to case. As our data tells us, in the aspect of for example psychiatry or abortion registry, confidentiality has a higher priority than it has in

many other cases. But in life threatening situations like for example a heart attack, availability and integrity have the higher priority over confidentiality.

4.2 Risk Management

It has become almost a given that every company or organization that use IT equipment run risk analysis for their software, network or other components connect to their systems. In the healthcare sector, risk and vulnerability analysis is vital in order to detect the risks with new implementations before the system is in use. It is important to have processes and strategies in place for the whole life cycle of the risk assessment – from how to conduct a risk assessment on a new system to how to deal with the issues and risks the risk assessment uncovers to how often a risk assessment should be put in motion. All the informants' states that a risk assessment is important to conduct for all parts of their systems and that they need continuous follow ups. However, based upon the findings in the interviews there is a slight difference in the procedures of how often a risk assessment is conducted from platform to platform. There is also some variety between the different health institutions.

ID5 has the following statement on risk assessment procedures for a larger platform in the health sector:

"We have run risk assessments of the platform every 1-2 months. We do not have very good tool support, so it has given us far too much work. Had it been up to me, we would have looked at more continuous risk assessments. That we should rather update, than start a new one every other month" (ID5).

This statement correlates with what the OAG found in their revision of the security in the Norwegian health sector. The OAG wrote in their report that:

"Hospital procurement and several of the informants at health organizations and regional ICT suppliers point out that reuse of Risk- and vulnerabilities analysis can be resource-saving, but this is done to a very small extent." (The Office of the Auditor General, 2021, p. 49).

The report goes on to further state that:

"Equally comprehensive Risk- and vulnerabilities analysis are often performed even if the system or equipment has already been used other health trusts. Informants refer to specific cases where this has been done even by a health trust bought identical medical technical equipment that they already had." (The Office of the Auditor General, 2021, p. 49).

What ID5 states of the continuity of the risk assessments differentiate a little with how ID4 talks about risk assessment in a smaller scale health application.

Here, the risk assessment is updating again since the last assessment was based upon what was known in May-June of 2021. As the informant states “*some things that were uncertain at the time are no longer uncertain and need to be treated a little differently*” (ID4). The complexity of the system itself may have an impact on how often a risk assessment is necessary for a specific platform service or application.

Continuity of risk assessment is something for example Amazon Web Services (AWS) and Microsoft underlines in their models for risk management. They both have models that are quite similar with some differences, but their goal is ultimately the same. Here there are some steps to follow when it comes to how to conduct risk assessments, and it should go in a continuous loop in order to make the systems as secure as possible (South, 2018; Microsoft, 2021). The OAG wrote in their audit of the Healthcare ICT systems from 2021 that the various health organizations in the different regions all perform risk- and vulnerabilities analysis when implementing new or changing existing systems. However, the report highlights that the analysis’ are not followed up systematically (Office of the Auditor General, 2021, p. 48).

ID6 goes more into detail on how the risk assessments are conducted. Although it is not directly linked to a specific platform or application, the informant states here how they in general conduct risk assessments.

“We have a scale of both consistency and probability. We use probability scale and consequence scale in addition to experience. Then you can say that we reuse risk elements between risk analyzes. If there are things we have encountered before, we will reuse them. The same applies to measures. Because it is unnecessary to reinvent the wheel all the way. But often things are used in a different way suddenly, and then the assessment may be different in terms of perhaps consequence and probability and then you have to reconsider. Reusing knowledge is something we focus on” (ID6).

Although there seems to be a slight variation in the process of when, how often and how to reuse the elements from the risk assessments, all of the informants agree on how to categorize the risks. All the informants agree that there is a combination of the probability and consequence that culminate in the total risk. There is a ranking system where the risks that are not acceptable are marked with red. If these risks exist, the system will not roll out. Risks that are acceptable but need to be followed up by measures are marked with yellow. The risks that are acceptable without any serious risks element to them are marked with green. This is a standard way for conducting a risk analysis. An example of such a model can be seen in figure 7 below (University of Bergen, 2021).

		Consequence				
		A Not hazardous	B A certain hazard	C Hazardous	D Critical	E Very critical
Probability/ Likelihood	5 Highly probable					
	4 Very probable					
	3 Probable		Event 1			
	2 Improbable			Event 2		
	1 Highly improbable					

Figure 7 Risk Matrix

4.3 Governance

Healthcare is possibly one of the most safety-critical and increasingly digitizing sectors, which is a trend we have seen over the past decade in Norway. Medical systems are becoming increasingly connected, exposing them to cybersecurity risks that could jeopardize patient health, safety, and privacy. Therefore, a well-established governance should be put in place at any organization within the healthcare sector to be able to follow both policies and procedures related to information security. Informant ID6 stated that:

“In general, governance in information security is important to maintain control over the different systems we have” (ID6)

Three of the informants we spoke to talked specifically about governance in healthcare, while the other informants spoke more about policies and procedures in general, without mentioning governance in terms of cyber security. The 3 informants who spoke about governance in terms of cyber security were all mentioning different policies they were following. ID6 and ID7 elaborated on how they used ISO27001 and ISO27002. ISO is a non-governmental international organization with 167-member national standards organizations. It brings professionals together to share information and establish voluntary, consensus-based, market-relevant International Standards that stimulate innovation and provide answers to global concerns through its members (ISO, 2022). During the

interviews, both ISO27001 and ISO27002 were mentioned. ISO27001 is an international standard that has been setting requirements for the establishment, implementation, maintenance, and continuous improvement of a management system for information security (Standard Norge, 2022). In addition to ISO27001, ISO27002 is often used as a supplementary standard that focuses on the information security controls that businesses may choose to deploy. These security controls are included in Annex A of ISO27001, which is what information security specialists frequently refer to when talking about information security controls. Unlike Annex A, which summarizes each control in one or two phrases, ISO27002 gives each control an average of one page. This is because the Standard describes how each control works, what its objective is, and how to put it into practice (IT Governance, 2021).

ID7 mention that they use ISO27001 as risk management and supply the management with security measures from both the Norwegian National Security Authority (NSM) and ISO27002.

“NSM has a lot of guidance that overlaps a little with ISO27001, but the basic principles are more to look at as measures than to look at as management. It fits together quite well, ISO27001 which says how to manage the area and then you have ISO27002 which is a place with lots of measures and NSM's basic principles are another place with measures” (ID7).

The NSM is, among other things, the Norwegian professional community for ICT security and the national alert and coordinating organization for cyber-attacks (Aspøy, 2022). NSM has therefore established some basic principles for ICT security. These principles are mainly targeted at organizations that deal with critical infrastructure, but they are adaptable to other organizations as well. The principles and measures here are meant to protect the information systems from damage, misuse, or unauthorized access (Norwegian National Security Authority, 2020, p. 5). The basic principles are split into four separate categories:

1. Identify & Map
2. Protect & Maintain
3. Detect
4. Handle & Restore

(Norwegian National Security Authority, 2020, p. 6)

Figure 8 below goes into detail on the different measures an organization can take within these four categories. They do not include every conceivable measure, but they do include what the NSM has found the most relevant for Norwegian businesses and organizations to protect their data. However, as the NSM states, these measures will help an organization to establish a good cyber-defense, but they cannot be seen as a guarantee that successful cyber-attacks can occur against an organization that has established these measures (Norwegian National Security Authority, 2020, p. 5).

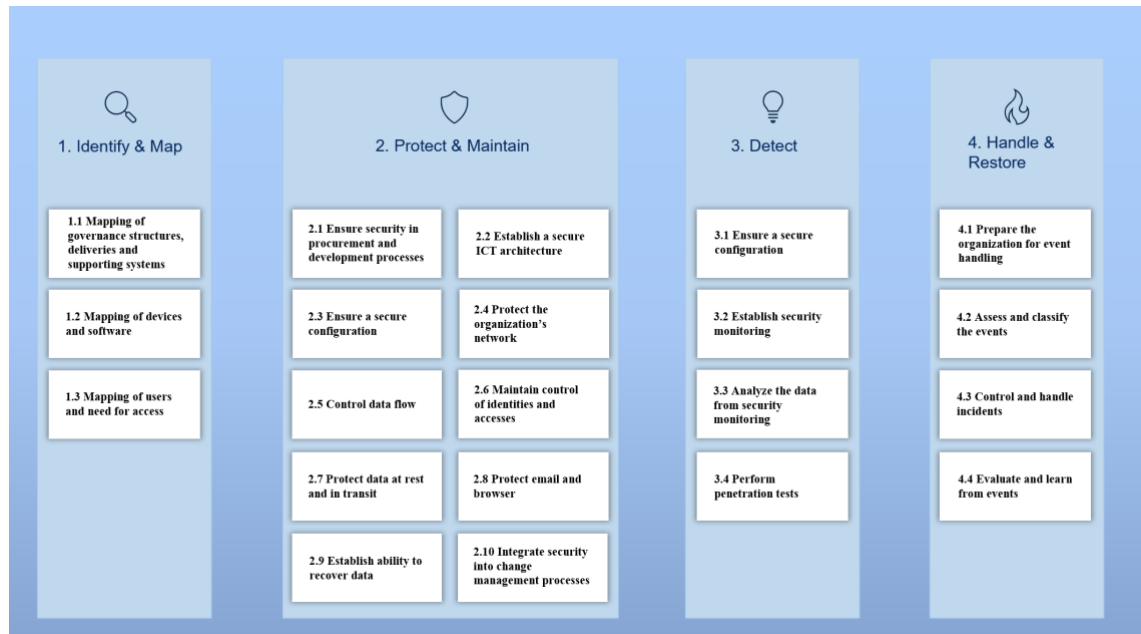


Figure 8 Model of The Norwegian National Security Authorities Basic Principles for ICT Security

ID6 has a slightly different perspective on NSM and ISO27001. The informant states that NSM are principles they are obligated to follow by law, while ISO27001 are a set of measures they follow to comply with NSM principles. When they implement new systems, or during patching of already existing systems, they use information security management systems (ISMS) for information security, which follows the ISO27001 guidelines.

“When there is value-chain risk, it is clear that risk assessments are central to this with the introduction of new systems. It's something that's done by default. In addition to all the requirements that must be met by those who deliver to us. So it takes a good number of rounds before things get in place. In the day-to-day operations, governance comes into the picture. How do we do it in relation to changes, then we have a change management process and security patching” (ID6).

The last informant (ID5) who spoke about governance elaborated more on risks and how they used governance systems to control these risks. Instead of mentioning principles they are obligated to follow by law (NSM), and principles such as ISO27001 and The National Institute of Standards and Technology (NIST) that may be used to be in accordance with the law, the informant mention “best practice” documents, and Microsoft templates for security.

“We use this type of Microsoft platform, which has very good templates and descriptions of how to do security in a very secure and structured way” (ID5).

When talking about security templates from Microsoft, we believe that the informant refers to different security programs that Microsoft offer its customers. Microsoft security is built on Zero Trust principles to protect data and provide access from anywhere, keeping the systems safe and productive. They offer several different security templates within hardware security, operating system security, user and identity security, cloud service security, privacy controls and security foundations (Microsoft, 2022).

The three informants who mentioned governance had slightly different opinions on why they use the security measures and risk management methods. The informant ID7 talked more about how helpful NSM and ISO27001 is while working with security in healthcare, while ID6 was more concerned with how they used ISMS to be in compliance with the law. ID5 elaborated more about how they used several different best practice documents and templates from Microsoft to follow the law. ID5 also confirmed that they use the templates and best practice documents while managing risks. OAG (2021) also stated in their audit that there is a need for risk management in data governance:

“In the regulations on electronic communication with and in the governance (eGovernment Regulations) § 15, emphasis is placed on establishing an internal control in the area of information security which is based on recognized standards for management systems for information security. The scope and structure must be adapted to risk management.” (Office of the Auditor General, 2021, p 18).

As explained before, the other 4 interview objects did not mention governance in terms of cyber security, but some of them did mention different security measures when it came to big data. ID1 elaborated about how they were corresponding with other departments about protective and detective measures, while ID3 were talking about how they were corresponding with EU members during the implementation of different projects such as the covid-19 certificate. Lastly, ID4 talked about how they used the lessons learned from different failures related to the implementation of different applications and systems over the past years. They have started to implement systems which are based on a more open way to develop, containing shared experiences across different health regions.

“We have learned a lot over the past decade, where many things happened in a very closed process where it was held internally and we published very little and got a lot of criticism for it, and it was a lot justified as well. Then there was something that was not justified.” (ID4).

4.4 Interoperability

During the interviews we discussed both the opportunities and limitations related to transfer of health records and other sensitive information within the Norwegian healthcare sector with the different informants. Two of the seven informants talked directly about issues related to sharing data between health regions in Norway. ID7 states that:

“Within security in healthcare today, the lack of sharing is the biggest weakness. And in relation to this, accessibility is the weak point” (ID7).

The three informants who speaks about this topic are all having the same opinion about the issues, where they are talking about how it is difficult to share and have control over data between the different health regions. Also, when talking about upcoming projects such as the “Health Analysis Platform”, “Health Platform” & “One Citizen, one Journal”, ID4 mention that they don’t share much information between the departments while developing it, while ID6 on the other hand elaborated about how they are cooperating/collaborating close between the different regions.

“We work closely together. In addition, we have a joint partner on information security called HelseCERT. They help us and they often arrange meetings together with the security leaders in the different regions to spread information and agree on joint measures and help each other” (ID6).

A Computer Emergency Response Team (CERT) is a collection of security experts working with handling computer incidents. In Norway, all CERTs are part of NSM (Norwegian National Security Authority, 2020). In the healthcare sector there are a CERT called HelseCERT. This CERT was established in 2011 and was meant to be a sector-specific response environment for the health sector. HelseCERT is located in Trondheim in the middle of Norway where they focus on cyber security where cyber is defined as all functions that are vulnerable through ICT, which is illustrated in the figure below with additional translation from Norwegian to English (Norwegian Health Network, 2022).

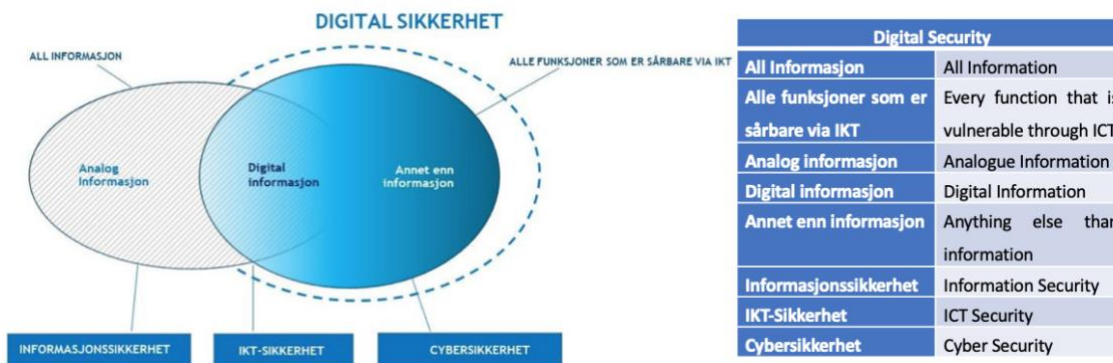


Figure 9 Norwegian HelseCERT

But ID6 also mentions that it is difficult to gather all information needed between the regions, because they use different operating systems, database structure etc.

“We have different computer systems and such, something is national and yea, another thing is that if we had just collected from other sources i.e., a large database and started going crazy with it, it would not take long before it becomes outdated either” (ID6).

Lastly, ID7 touch upon some interesting information where the informant elaborated around how they share information between the different health regions in urgent situations. The informant told us about how the doctors and nurses are forced to use fax machines to transfer EHR when a patient is in a hospital outside of his or her registered health region.

“There is still a lot of use of fax machines, that you have to call from one hospital to another and ask if you can get a fax about that patient to see the journal. There are employees who have never encountered a fax before, and that should not be the case. There are safer ways to get the data from the recipient. There we have a project, there is the sharing of journals between companies, but it is not so widespread yet” (ID7).

When we asked this informant to further elaborate about the situation with fax machines and transfer of health data, the informant further described the situation:

“It is at least at the expense of availability. It's obvious. Probably integrity also because you have to move and fax and then it is easy for employees to suddenly send the wrong document or something like that. You call and ask for a person and it was the name-brother/sister you got instead. If there are manual routines, it is easy to forget things. Integrity is probably worse as we have it now than if it were collected. Confidentiality then it is more unclear” (ID7).

In fact, Pérez (2019) states that fax machines are not in compliance to the GDPR because Fax machines are rarely encrypted, but they can keep electronic copies of

papers delivered or received. As a result, any sensitive data sent can end up on an insecure hard drive. Only when they are overwritten by subsequent papers are they erased. This can lead to a second issue, which is the fax server. Most in-house servers do not have encryption software installed. Like fax machines, these servers can keep insecure electronic copies of papers for a long time. To make matters worse, when a server's capacity is reached, corporations frequently print paper copies to store them, making them even more exposed and increasing the possibility of non-compliance. The last vulnerability that Pérez (2019) mention is that unless you wait for the fax to arrive, there's a good chance that someone else will see important papers. Most businesses have shared fax machines that anybody can use, which is in violation of legislation such as GDPR, MIFID II, HIPAA, and FERPA (Pérez, 2019).

The other informants who did not speak about the transfer of EHR itself, but more about communication within the healthcare sector in Norway did mention some room for improvements. ID3 spoke about how there, by the informant's personal experience, has been a lack of communication and transparency when something goes wrong in terms of information security. The informant further elaborates how they are working on improving the communication between the different regions and the HelseCERT.

"If you go to "Helsenorge" (application for patients in Norway) and ask about your journal, then we have to talk to other health regions, to find where the data is, we do not have a copy of all the data that is out there, so then we discuss risk, and try to see the whole, so then we get a much better picture of what has happened. Then we get our experiences and the experiences out there, so if we are going to talk about things, then where there is openness and more common dialogue around what has happened, and what measures have been used / implemented" (ID3).

When we look at our data and the different opinions related to interoperability in the healthcare sector, we can see that there is a huge difference within the different systems. The data tells us that HelseCERT is a well-developed Computer Emergency Response Team which has a mature interoperability level when working with security across the health regions. However, when it comes to the systems themselves, like the different health journals and other platforms or systems used in the different regions, the interoperability level is much lower. As the data shows, it is not as easy to share patient records between regions and between private and public medical facilities as one should think.

5 DISCUSSION

This chapter will provide a discussion of some of the most essential empirical findings and the previous studies found in the systematic literature review. The chapter will also explain why the data being presented is important in the context and how it might be interpreted in the field of research.

5.1 Privacy

As described in findings, privacy plays a big role in healthcare. All the informants mentioned privacy in different context, and how important it was to focus on it during the interview, which is coherent with the previous studies, which stated that in 2016, 80 percent of large-scale enterprises faced severe privacy risks connected to big data (Al-Shomrani et al., 2017).

The reason for privacy playing a big role in healthcare is due to the fact that EHR contains highly sensitive information. As the Norwegian healthcare sector are developing new platforms to share health data, there is a need for changes in their systems. As Jain et al (2016) describes, patient data is initially stored in data warehouse's with various levels of data protection, and in the case of big data in healthcare, security measures used from the old systems are ineffective when dealing with data sets that are instinctually heterogenous.

Our data proves that anonymization of data is important, and it explains how it is used to be in compliance with regulations and laws such as GDPR. Cavanillas et al (2016) on the other hand, argued that there can be problems related to anonymizing data, especially when different data sets may be combined to find connections between the data. The data states that there is a disagreement about anonymization. Concerns about the identifying individuals due to a small number of citizens in different regions, the degree of sensitivity of the data, and the availability of the data were all elaborated in the interviews.

Patil & Seshadri (2014) argued that security and privacy is connected to data more than certifications and policies. One of the informants described how privacy and security are connected to each other, where there is good information security that ensures good measures on patient privacy. The main difference between previous studies and the informants is that the informants were more divided in the importance of privacy in healthcare, mostly due to type of health data, the use area, and the urgency of patient care in certain situations. While the previous

studies focus more on general concerns related to patient privacy without any specific situation in mind, this study has shown that the issue regarding privacy differs from one situation to another.

5.2 Risk Management

As established in chapter 4.2, the OAG found that the various health organizations in the different regions all perform risk- and vulnerabilities analysis when implementing new or changing existing systems. So how does this audit from 2021 compare our data? The informants said that there is systematical follow-up of risk assessments in the health sector. However, there seems to be a variety on how often the risk- and vulnerabilities analysis are conducted from one platform to another. On a smaller scale project, the informant told us that a new risk assessment was done approximately a year after the last one. On a large-scale project another informant told us that a risk assessment was done every other month.

One of the issues that emerged during the study was the lack of reusing risk assessments. One of the informants was specific in that it would have been easier to update and run continuous risk assessments instead of starting a new one every other month. This was also one of the issues that the OAG mentioned in their report when they reviewed the ICT security in the Norwegian health sector. One of the problems that emerged through this research was the lack of sharing risk- and vulnerabilities assessments from one health region to another. One health region has to do their own risk assessment even though the same equipment is already in use in another region where it has been risk assessed. Figure 10 below illustrates this situation with Health Region A and Health Region B based on our own data collected.

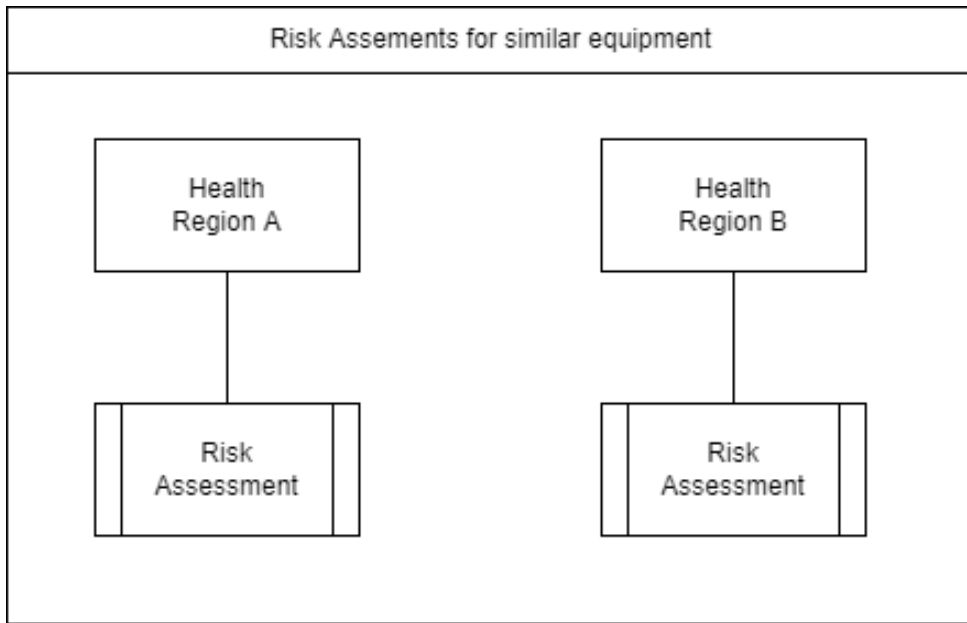


Figure 10 Risk Assessments for Similar Equipment

It takes a lot of resources to conduct a risk assessment. OAG (2021) estimate that it takes between 150-200 hours for risk assessments for different ICT systems and equipment. As both informants from this research and informants in the OAG report says, there is a desire to have some changes in the risk assessment routines. Figure 11 illustrates a concept on how we propose a sharing of risk assessments across the health regions can be done.

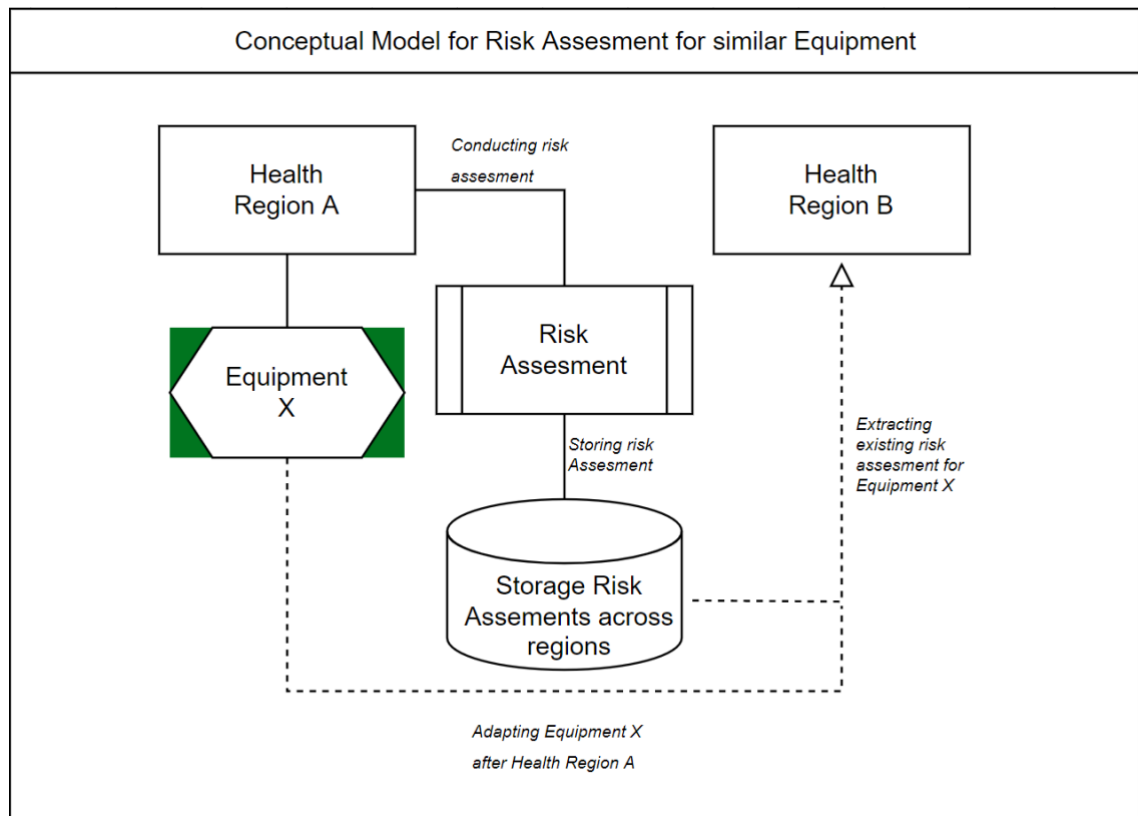


Figure 11 Conceptual Model for Risk Assessment for Similar Equipment

In this model, both Health Region A and B shall adapt the same equipment. Health Region A adapts it first and therefore has to conduct a risk assessment. This risk assessment is then stored in a database that is accessible for every health region. Health Region B then decides to adapt the same equipment and extract the risk assessment conducted by Health Region A in the process.

Sharing and reusing risk assessments as explained in figure 11 can provide more efficiency for the health regions. However, there is important to know how to risk assess in an efficient and proper way. During this research the informants said that they risk assess every part of the systems they use. They also look at security controls to be implemented and that they monitor their systems during their lifetime. This correlates well with findings from documents that has been analyzed from AWS and Microsoft. AWS has a six-step model for their risk management, and Microsoft has a four-step model. These models are quite similar to one another, however, there are some minor details that differentiate them. Since the healthcare sector deals with highly sensitive PII, combining these to create a seven-step model seems like the most secure option.

The seven steps in the model are as follows:

1. Categorize IS
2. Identify Risks
3. Select Security Controls
4. Implement the Controls
5. Assess Security Controls
6. Authorize System
7. Monitor

The health sector is already categorizing their information systems. This is necessary to maintain control over their IT structure. The next two steps are to identify risks and selection of controls. Here, there should be a cooperation between the health regions through HelseCERT. HelseCERT can assist in providing information of which risks are out there and how one can limit the risks with the help of controls. The controls will be selected from existing frameworks like ISO or similar standards. NSM security measures can also here help to limit the risks. The next step is to implement these controls and assess them along the way. When all these steps are concluded, and the risks are limited to no longer be in the red zone (figure 7 in chapter 4.2) the system can be authorized and used. When the system is operating it is important to monitor the system along the way. New risks can occur, and better controls can emerge as technology develops and therefore it is the recommendation to have all these steps in a continuous loop throughout the lifespan of the system. The concept of this model is illustrated in figure 12 below.

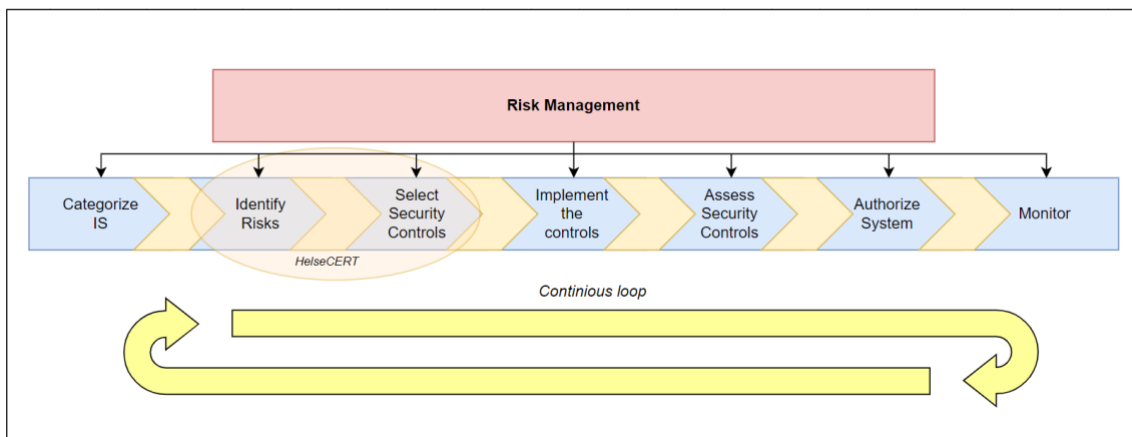


Figure 12 Risk Management

This research has collected data that has established two main issues related to risk management; sharing risk assessments across the regions and how to conduct an effective risk assessment with the use of a sector specific risk management framework. Figure 11 represents our own contribution for sharing risk assessments across the different health regions. Figure 12 represents our contribution to risk

management within the healthcare sector based on our findings. We believe that following the 7 steps identified here will strengthen the security of health systems in the Norwegian Healthcare Sector.

5.3 Information Security Governance

As stated in the literature review, a successful data governance program fosters the development of formal data standards and norms, as well as data supervision, so that decision makers have access to high-quality, consistent, and timely data to address the issues and opportunities facing the healthcare organization (Trom & Cronje, 2019, p. 649).

As described in the findings, the informants were a bit divided when it came to governance. Some of the informants mentioned data governance specifically, while the others talked more about governance in general. In our literature review, Alofaysan et al. (2014) and Trom & Cronje (2019) both argues that a good governance is focused more on people and procedures, rather than technologies, which can reflect the response we got from the informants. The informants were mentioning ISO and NSM as the most important parts of their data governance, either if it was to use to be in accordance with the law or for their own systems to be as secure as possible. If we look at the table 5 in chapter 2.2.3 about data governance in healthcare and combine it with the information collected during the interviews, we can see that the Norwegian Healthcare sector is based on an organizational-level Governance. This means that the key dimensions of their data governance should contain *“Data domains”, “Governance goals, Governance forms and “Value/application”*. This is due to the fact that organizational-level governance is described as where a hospital or other clinical organizations is the source and major consumer of PHI data generated on its own systems (EHR). Individual clinical data, as well as operational data such as records of clinician-provided services and financial payments, are kept on file at the hospital (data domains). Government requirements may compel data to be shared with accredited people, researchers, or patients, and this data are owned by the healthcare organization (governance goals and forms). The organization is also in charge of ensuring patient privacy (as required by law) and data security, as well as making data easily accessible to clinicians as they perform their duties and analysts to assess the efficiency and quality of service delivery (governance goals, data value), primarily through ICT such as EHR systems or data repositories (governance forms) (Winter & Davidson. 2019, p 39).

The answers we got from the informants stated that they were all referring to principles such as ISO27001 and ISO27002 to be compliant with NSM and laws, which then again would make their systems secure. But as Trom & Cronje (2019)

mention, there is a need for supervision as well as following principles and norms. Not only the literature states that there is a need for supervision in governance.

As we stated in 5.2, there is a lack of follow-up when it comes to risk management within the health sector in Norway, which would affect the supervision when it comes to data governance. We don't believe that the health organizations in Norway would be able to maintain a good data governance when it is only based on principles and regulation rather than combining it with risk assessment methods that all regions should take part of. Risk management procedures combined with the use of formal data standards and norms would help the decision makers have access to high-quality, consistent, and timely data to address the issues and opportunities facing the healthcare organization, but for now, it would be almost impossible to have control over the different issues and risks that may occur. Figure 13 below illustrates the different categories of Organizational-level Information Security Governance we find necessary. The circles illustrate the data governance dimensions the categories represents. Laws & Regulations and Standards/measures ensure the governance goals and forms, while IT-policies, organizational policies and Access Control Policies ensure that the data maintain its value to both organizations and patients.

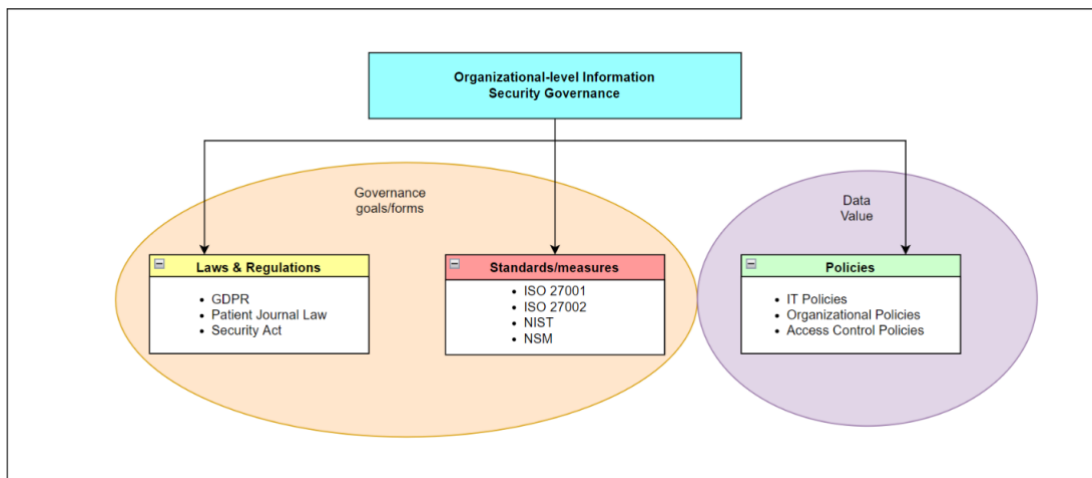


Figure 13 Organizational-level Information Security Governance

Figure 13 is based on our findings and is our own contribution to what we believe that an Organizational-level Information Security Governance should contain.

5.4 Interoperability Frameworks

In the findings we can see that the informants were a bit divided in their opinion of interoperability in the Norwegian healthcare sector. While some of the

informants elaborated about how good helseCERT works for them, and how they communicate across the different regions, other stated that there is a huge lack of interoperability in many systems and procedures. As OAG (2021) report stated, there is a lack of interoperability between the different health regions. Based on table 6 from 2.2.4, which describes different levels of interoperability in the healthcare sector, we can see that the Norwegian healthcare sector can be found somewhere between level-0 and level-1 interoperability. Level-0 interoperability is where a system is stand-alone and has no preconditions to be shared with another. Level-1 of interoperability, the adoption of a communication protocol for data transmission across systems is required. Due to the fact that the informant elaborated about how they have to share EHR with sensitive information on Fax-machines, we place the current system for sharing data at level-0 interoperability. The figure below illustrates the current state of interoperability in the Norwegian Healthcare based on our findings.

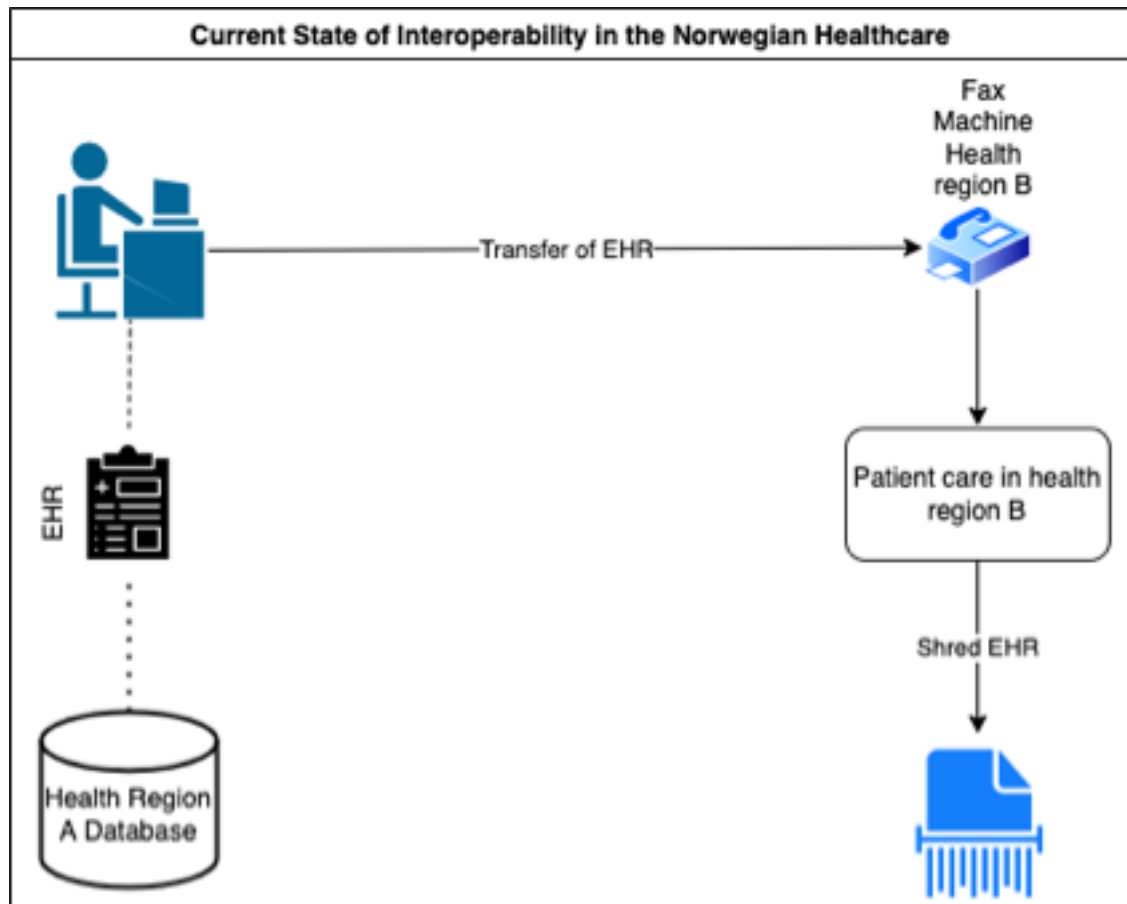


Figure 14 Current State of Interoperability in the Norwegian Healthcare

Figure 14 illustrates how the workflow is if a patient need care in health region B, while the EHR is stored in health region A. Doctors/nurses has to call to health

region A to gain access to the health journal stored in region A, and the transfer is done by fax-machines.

To avoid using fax machines or other vulnerable methods for sharing health data between the different regions in Norway, the different health regions implement a more suitable interoperability level. As the current system is at a level-0 interoperability, there is a significant need for improvement. We believe that level-3 or semantic interoperability is the level that the Norwegian Healthcare Sector should implement. Semantic interoperability refers to the ability of two or more systems to automatically grasp the information transmitted in a meaningful and accurate manner in order to produce valuable outcomes as specified by the systems' end users. The figure below illustrates how we believe that this level should look like in the Norwegian Healthcare Sector.

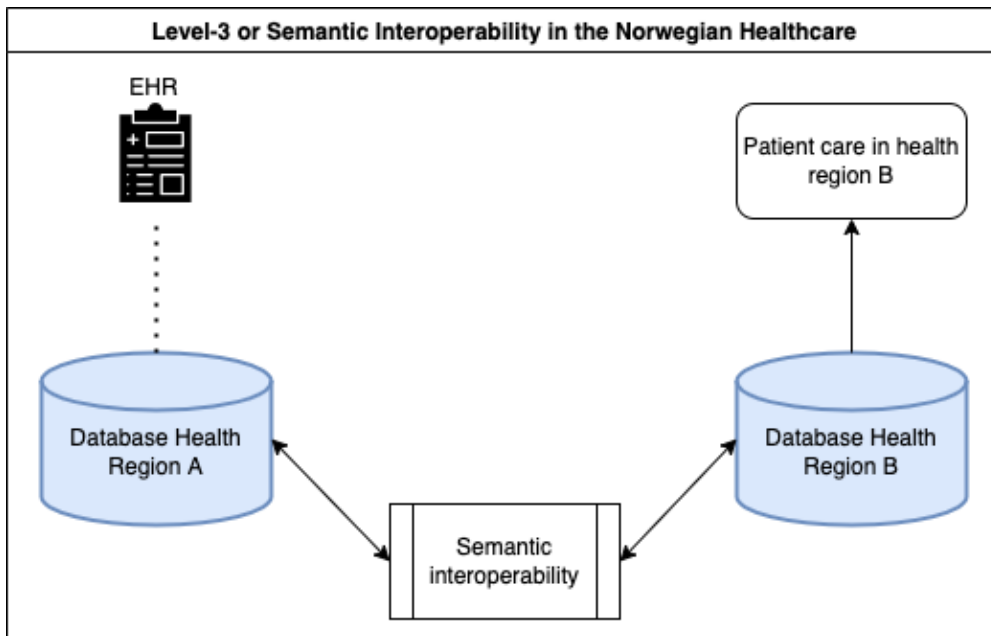


Figure 15 Level-3 or Semantic Interoperability in the Norwegian Healthcare

Figure 15 states the difference when a patient from region A needs care in region B. While figure 14 shows how doctors/nurses must call hospitals to get access to health journals through fax-machines, figure 15 visualizes how a semantic interoperability system will help both the security of the sensitive information and share accurately information. This will also save both time and money for the health regions.

5.5 Final Result

This chapter has discussed the findings from the literature review, interviews, and document analysis. But do they help to find an answer to the research questions and the research problem? The first research question proposed in chapter 1 was as follows:

RQ1:

- How can data governance framework mitigate security risks in the healthcare sector?

This research proved that security governance and risk management are closely linked together. If there is a systematic framework for how to have good risk management and good governance in the organization. This culminated in figure 12 (risk management) and figure 13 (Organizational-level information Security Governance). But to answer the question it is important to see these in context with one another. Figure 16 below put these models together to see how the Norwegian health sector can use governance and risk management together to achieve compliance for their system. In this study, that system is an interoperability system that gathers a large amount of EHR. When compliance is ensured, that system can be operational. As with the risk management model, it is important to run a continuous loop of this framework. For instance, a law can be implemented or change (as seen with Schrems II and the GDPR) and new requirements for the system will thereby emerge. Then one has to run new assessment in order to make the system compliant again. Therefore, a continuous loop of this model is necessary in order to achieve the best level of security the health sector want for their system.

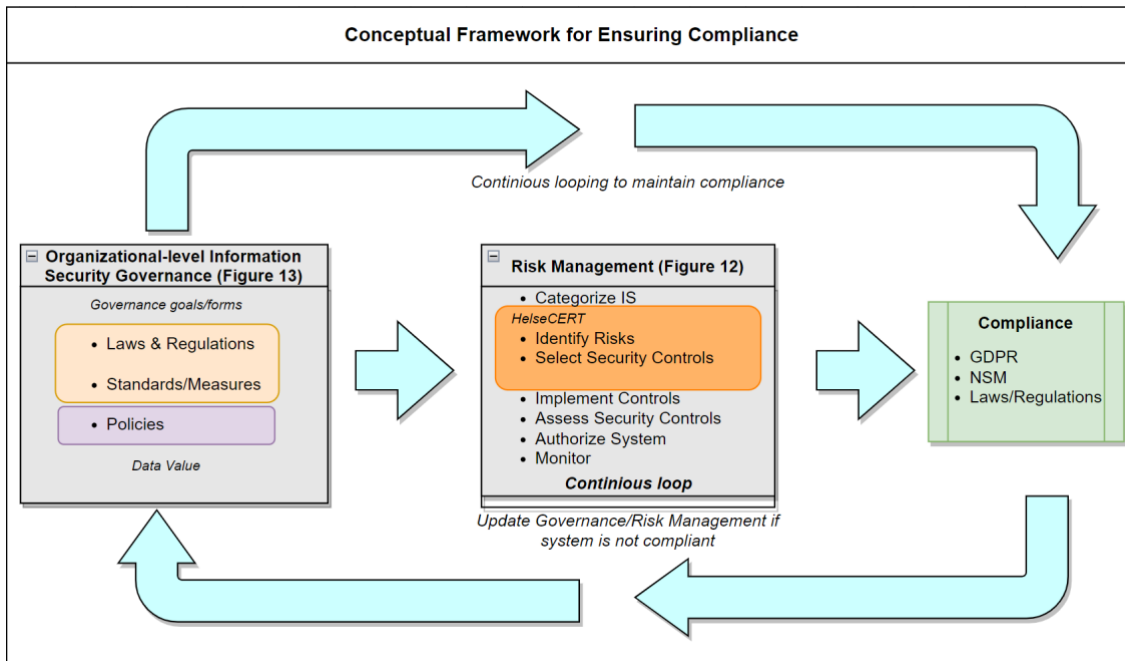


Figure 16 Conceptual Framework for Ensuring Compliance

RQ2:

- How can Norwegian the Norwegian healthcare sector attain interoperability in integrated systems?

According to the systematic literature review, the development of widely accepted standards can aid in the resolution of interoperability. As there are several different health regions in the Norwegian Healthcare Sector, with additional internal systems for storing EHR, the need for an integration of interoperability is significant. The (EIF) may be a base for these health regions to use. This framework enables health organizations to collaborate toward mutually beneficial goals, incorporating the sharing of information and expertise between them via the business processes they support and data exchange between their ICT systems. This framework is based on a semantic interoperability, which is described in figure 15 in chapter 5.4.

By applying the "Conceptual Framework for Ensuring Compliance" (figure 16), we believe that the health organizations can improve their current (and upcoming) systems, because it will increase compliance through continuous risk management procedures as well as being in accordance with the law and different policies conducting organizational-structured Information Security Governance. As our data told us in chapter 4.4, HelseCERT should be included when it comes to interoperability between different regions in the Norwegian Healthcare Sector. If the different health regions implement a semantic interoperability (see figure 15), which includes a cooperation between HelseCERT, the different health regions, and other health institutions such as smaller private clinics, we believe that it will strengthen the interoperability of their systems, and improve storing, processing,

and transferring of Electronic Health Records. Thus, together with the framework represented in figure 16, it can contribute to a solution of the research problem for this study:

How can healthcare organizations ensure security in the Big Data era?

6 CONCLUSION

The Norwegian healthcare sector got a damning review from the report of the Office of the Auditor General when it comes to cyber security. They were able to take control of three out of four health regions' ICT infrastructure. The report also proved that there were flaws in the security culture and lack of knowledge of information security among personnel and management in the health sector. With the rollout of new interoperable systems like Helseplattformen and Helseanalyseplattformen, there is a need to close the gaps between the risks and security in healthcare ICT systems. This research is based on literature review from academic journals and publications, and empirical qualitative studies through interviews and document analysis that together has laid the foundation for this thesis.

This thesis has taken a dive into how security in the healthcare sector can be managed from an information security governance and risk management perspective. There are a lot of issues regarding the healthcare ICT systems. Strict laws, old systems and several different health regions plays it part in making Information Security and Interoperability in the healthcare a complex challenge. During this research, the goal has been to establish answers to some of these issues. Answering the research problem and the research questions has been the goal of this research, and chapter 5 states the answers of this research.

This thesis has provided answers on how the Norwegian healthcare sector can govern and manage information security across the different regions. By following the frameworks presented in chapter 5.2, 5.3 and 5.4 which adapts to the structure of the Norwegian healthcare system, we believe that it will be easier to adapt level-3 interoperable systems like "Helseanalyseplattformen" and "Helseplattformen" in way that meets security standards in the current ICT landscape.

7 LIMITATIONS

This section goes through the various limitations that have been faced and how they may have influenced the project.

7.1 Research Problem

Looking at the research problem we can see that it may be a bit broad for a master thesis. There are a lot of aspects to cover while working with security in Big Data platforms, especially in the healthcare sector. The thesis focuses mainly on risk management, security governance and interoperability, which might increase the usefulness of the thesis. However, we believe that the additional research questions would help narrow down the aspects of Big Data in healthcare. Some parts of security in Big Data could have been added, such as technical parts, but the intended goal was to answer the research questions from a management perspective rather than technological.

7.2 Interviews

Due to the fact that the interview objects were located all around Norway in the different health regions, it was not possible for us to conduct physical interviews. The solution to this was to perform the interviews online. Some research indicates that performing interview physically rather than digitally would result in better responses from the informants. However, since we are all used to online meetings during the pandemic, we believe that conducting the interviews digitally had no significant impact on the responses we received.

Another potential limitation is that, due to the sensitive nature of the planned questions, it may be difficult to obtain correct and detailed replies to interview questions. Especially since the informants we interviewed were all either in charge of the security at their department or in a team who was. Organizations may be hesitant to share information with students, particularly if it concerns their internal security. This issue was resolved by signing a non-disclosure agreement with the client organization, which stipulated that sensitive information would not be published in the thesis and that interview subjects and their organizations would remain anonymous.

7.3 Sample

During the interviews we did not meet any problems with saturation because we interviewed informants from all the different health regions. Conducting more interviews could enrich the study by developing the interview guide to narrow the research. We still find 7 informants satisfying for this study as we interviewed different employees with security management responsibilities across different health regions and health organizations. In addition, we tried to get in touch with more people to interview; some of them did not respond and some told us that they did not find their position at their workplace relevant for the study. This may potentially have given us a lack of in knowledge of the study area as they might have provided some additional information that could be of importance for this research.

To compliment the 7 informants for this study we did an additional document analysis to get a broader perspective of the study area. This allowed us to either verify or create doubt in the answers and compare them to the findings of the systematic literature review.

7.4 Time Constraints

Due to the project's tight timeline, we had to divide our efforts across several aspects of the project. Selecting interview subjects was one of the most time-consuming and resource-intensive operations. Additionally, numerous hours of work were lost because some potential interviewees did not respond to our invitation to participate in our research study.

During the middle of the project period, we received some feedback suggesting conducting the literature review differently. This led to a set back at the time, since we had to re-write almost the whole thesis. But, looking at it now, we're thankful we were told to do these changes because this project has turned out lot better than it would without that feedback.

8 FUTURE RESEARCH

In this thesis, several different security aspects of Big Data in Healthcare have been discussed. Implementation of risk management, security governance and interoperability systems has been elaborated about. Further research should include more in-depth analysis of standards such as ISO and NIST, roles when it comes to security governance, and more technical aspects of a semantic interoperability system. If we look at the research questions from this thesis, we believe that it could be interesting to do more research on the different systems in the Norwegian Healthcare Sector from a technical point of view. It could then be possible to implement our conceptual frameworks within a semantic interoperability system. Our contribution based on managerial perspective being implemented in a more technical framework to fulfill the need of security when it comes to the transfer of sensitive data in the health regions in Norway.

It could also be interesting to get additional qualitative data in the healthcare sector based on security awareness and routines among different employees in the different health regions. We assume this because we do not believe that effective data governance can be achieved without considering all areas of ICT. From a broader international perspective, a generalization of different health institutions from different international regions can be interesting to look at. This might support exploring similarities among countries and across borders.

RESOURCES

- Abouelmehdi, K., Beni-Hssane, A., Khaloufi, H., & Saadi, M. (2017). Big data security and privacy in healthcare: a review. *Procedia Computer Science*, 113, 73-80. <https://doi.org/10.1016/j.procs.2017.08.292>
- Alofaysan, S., Alhaqbani, B., Alseghayyir, R., & Omar, M. (2014, March). The significance of data governance in healthcare. In *Proceedings of the International Joint Conference on Biomedical Engineering Systems and Technologies* (Vol. 5, pp. 178-187). <https://doi.org/10.5220/0004738101780187>
- Al-Shomrani, A., Fathy, F., & Jambi, K. (2017, March). Policy enforcement for big data security. In *2017 2nd international conference on anti-cyber crimes (icacc)* (pp. 70-74). IEEE. <https://doi.org/10.1109/Anti-Cybercrime.2017.7905266>
- Amaratunga, D., Baldry, D., Sarshar, M., & Newton, R. (2002). Quantitative and qualitative research in the built environment: application of “mixed” research approach. *Work study*.
- Aspøy, A. (2022). *Nasjonal Sikkerhetsmyndighet*. From: https://snl.no/Nasjonal_sikkerhetsmyndighet
- Baro, E., Degoul, S., Beuscart, R., & Chazard, E. (2015). Toward a literature-driven definition of big data in healthcare. *BioMed research international*, 2015.
- Befring, A & Sand, I-J. (2020). *Kunstig intelligens og big data i helsesektoren*. Gyldendal.
- Binjubeir, M., Ahmed, A. A., Ismail, M. A. B., Sadiq, A. S., & Khan, M. K. (2019). Comprehensive survey on big data privacy protection. *IEEE Access*, 8, 20067-20079. <https://doi.org/10.1109/access.2019.2962368>
- Cavanillas, J. M., Curry, E., & Wahlster, W. (2016). New horizons for a data-driven economy: a roadmap for usage and exploitation of big data in Europe. Springer Nature. <https://doi.org/10.1007/978-3-319-21569-3>
- Chauhan, R. S. (2019). Unstructured interviews: are they really all that bad?. *Human Resource Development International*, 1-14. <https://doi.org/10.1080/13678868.2019.1603019>
- Cowie, M. R., Blomster, J. I., Curtis, L. H., Duclaux, S., Ford, I., Fritz, F., ... & Zalewski, A. (2017). Electronic health records to facilitate clinical research. *Clinical Research in Cardiology*, 106(1), 1-9. <https://doi.org/10.1007/s00392-016-1025-6>
- Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: management, analysis and future prospects. *Journal of Big Data*, 6(1), 1-25. <https://doi.org/10.1186/s40537-019-0217-0>

- Denardo, A. M. (2002). Using NVivo to analyze qualitative data. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.83.5090>
- Gisle, J. (2018, November. 13). *Personvernforordningen*. From: <https://snl.no/Personvernforordningen>
- Gregar, J. (1994). Research Design (Qualitative, Quantitative and Mixed Methods Approaches). *Book published by SAGE Publications*, 228.
- Gross, J.M.S., (2018). Document Analysis. *SAGE Publications Inc*. <https://dx.doi.org/10.4135/9781506326139.n209>
- Gupta, S., Kar, A. K., Baabdullah, A., & Al-Khowaiter, W. A. (2018). Big data with cognitive computing: A review for the future. *International Journal of Information Management*, 42, 78-89. <https://doi.org/10.1016/j.ijinfomgt.2018.06.005>
- Helseplattformen.no (2022, February. 2). *Hva er helseplattformen?* From: <https://helseplattformen.no/om-oss/generelt-sporsmal-og-svar#hva-er-helseplattformen>
- Hemingway, H., Asselbergs, F. W., Danesh, J., Dobson, R., Maniadakis, N., Maggioni, A., ... & Innovative Medicines Initiative 2nd programme, Big Data for Better Outcomes, BigData@ Heart Consortium of 20 academic and industry partners including ESC. (2018). Big data from electronic health records for early and late translational cardiovascular research: challenges and potential. *European heart journal*, 39(16), 1481-1495. <https://doi.org/10.1093/eurheartj/ehx487>
- Hennink, M., Hutter, I., & Bailey, A. (2020). *Qualitative research methods*. Sage.
- Iroju, O., Soriyan, A., Gambo, I., & Olaleke, J. (2013). Interoperability in healthcare: benefits, challenges and resolutions. *International Journal of Innovation and Applied Studies*, 3(1), 262-270.
- ISO. (2022). *About us*. From: <https://www.iso.org/about-us.html>
- IT Governance. (2021, July. 22). ISO 27001 vs. ISO 27002: What's the difference? From: <https://www.itgovernance.co.uk/blog/understanding-the-differences-between-iso-27001-and-iso-27002>
- Jaïdi, F., Labbene-Ayachi, F., & Bouhoula, A. (2016). Advanced techniques for deploying reliable and efficient access control: Application to E-healthcare. *Journal of medical systems*, 40(12), 1-9. <https://doi.org/10.1007/s10916-016-0630-2>
- Jain, P., Gyanchandani, M., & Khare, N. (2016). Big data privacy: a technological perspective and review. *Journal of Big Data*, 3(1), 1-25. <https://doi.org/10.1186/s40537-016-0059-y>
- Juddoo, S., George, C., Duquenoy, P., & Windridge, D. (2018). Data governance in the health industry: Investigating data quality dimensions within a big data context. *Applied System Innovation*, 1(4), 43. <https://doi.org/10.3390/asi1040043>

- Kouroubali, A., & Katehakis, D. G. (2019). The new European interoperability framework as a facilitator of digital transformation for citizen empowerment. *Journal of biomedical informatics*, 94, 103166. <https://doi.org/10.1016/j.jbi.2019.103166>
- Lakshen, G. A., Vraneš, S., & Janev, V. (2016, November). Big data and quality: A literature review. In *2016 24th telecommunications forum (TELFOR)* (pp. 1-4). IEEE.
- Le Bris, A., & El Asri, W. (2016). State of cybersecurity & cyber threats in healthcare organizations. *ESSEC Business School*, 12.
- Leung, L. (2015). Validity, reliability, and generalizability in qualitative research. *Journal of family medicine and primary care*, 4(3), 324. <https://doi.org/10.4103/2249-4863.161306>
- Magaldi, D., & Berler, M. (2020). Semi-structured interviews. *Encyclopedia of personality and individual differences*, 4825-4830.
- Microsoft. (2021). *Microsoft 365 Risk Management program*. Retrieved from <https://docs.microsoft.com/en-us/compliance/assurance/assurance-risk-management-program>
- Microsoft. (2022). *Windows Security*. From: <https://docs.microsoft.com/en-us/windows/security/>
- Morabito, V. (2015). Big data governance. *Big data and analytics*, 83-104. https://doi.org/10.1007/978-3-319-10665-6_5
- Myers, M. D., & Newman, M. (2007). The qualitative interview in IS research: Examining the craft. *Information and organization*, 17(1), 2-26. <https://doi.org/10.1016/j.infoandorg.2006.11.001>
- Norwegian Health Network. (2022). *HelseCERT*. From: <https://www.nhn.no/Personvern-og-informasjonsikkerhet/helsecert>
- Norwegian National Security Authority. (2020, June. 26). *Event Management*. From: <https://nsm.no/fagomrader/digital-sikkerhet/nasjonalt-cybersikkerhetssenter/handtering-av-dataangrep/hendelseshandtering>
- Okoli, C., & Schabram, K. (2010). A guide to conducting a systematic literature review of information systems research.
- Patil, H. K., & Seshadri, R. (2014, June). Big data security and privacy issues in healthcare. In *2014 IEEE international congress on big data* (pp. 762-765). IEEE. <https://doi.org/10.1109/BigData.Congress.2014.112>
- Patton, M. Q. (2014). *Qualitative research & evaluation methods: Integrating theory and practice*. Sage publications.
- Pérez. (2019). 3 security risks of using Fax machines and how to avoid them. From: <https://www.cloudworldwideservices.com/en/3-security-risks-of-using-analog-fax-machines-and-how-to-avoid-them/amp/>
- Salas-Vega, S., Haimann, A., & Mossialos, E. (2015). Big data and health care: challenges and opportunities for coordinated policy development in the EU. *Health Systems & Reform*, 1(4), 285-300. <https://doi.org/10.1080/23288604.2015.1091538>
- Schwarz, G. M., & Stensaker, I. G. (2016). Showcasing phenomenon-driven research on organizational change. *Journal of Change Management*, 16(4), 245-264. <https://doi.org/10.1080/14697017.2016.1230931>

- Seh, A. H., Zarour, M., Alenezi, M., Sarkar, A. K., Agrawal, A., Kumar, R., & Ahmad Khan, R. (2020, June). Healthcare data breaches: Insights and implications. In *Healthcare* (Vol. 8, No. 2, p. 133). Multidisciplinary Digital Publishing Institute. <https://doi.org/10.3390/healthcare8020133>
- Sessler, D. I. (2014). Big Data—and its contributions to peri-operative medicine. *Anaesthesia*, 69(2), 100-105.
- South, M. (2018). *Scaling a governance, risk and compliance program for the cloud, emerging technologies, and innovation*. Retrieved from <https://aws.amazon.com/blogs/security/scaling-a-governance-risk-and-compliance-program-for-the-cloud/>
- Standard Norge. (2022, April. 27). NS-EN ISO/IEC 27001 Ledelsessystemer for informasjonssikkerhet. From: <https://www.standard.no/fagomrader/ikt/it-sikkerhet/isoiec-27001/>
- The Directorate of eHealth. (2019, June. 25). Overordnet IKT-ROS for helse- og omsorgssektoren. In The Directorate of eHealth. Retrieved February 2. 2022 from: <https://www.ehelse.no/publikasjoner/overordnet-risiko-og-sarbarhetsvurdering-for-ikt-i-helse-og-omsorgssektoren>
- The Directorate of eHealth. (2021, October. 13). *Helsedataservice*. From: <https://www.ehelse.no/programmer/helsedataprogrammet/helsedataservice>
- The Directorate of eHealth. (2021, December. 22). *Helseanalyseplattformen*. From: <https://www.ehelse.no/programmer/helsedataprogrammet/helseanalyseplattformen>
- The Norwegian Government. (2019, October. 30). *Ny personopplysningslov*. From: <https://www.regjeringen.no/no/tema/statlig-forvaltning/personvern/ny-personopplysningslov/id2340094/>
- The Norwegian National Security Authority. (2020). *NSM's grunnprinsipper for IKT-sikkerhet*. Retrieved from <https://nsm.no/getfile.php/133735-1592917067/Filer/Dokumenter/Veiledere/nsms-grunnprinsipper-for-ikt-sikkerhet-v2.0.pdf>
- The Office of the Auditor General. (2022). *About us*. <https://www.riksrevisjonen.no/en/about-the-oag/about-us/>
- The Office of the Auditor General. (2021). Riksrevisjonens undersøkelse av helseforetakenes forebygging av angrep mot sine IKT-systemer. In *The office of the Auditor General*. Retrieved February 2. 2022 from <https://www.riksrevisjonen.no/globalassets/rapporter/no-2020-2021/undersokelse-av-helseforetakenes-forebygging-av-angrep-mot-sine-ikt-systemer.pdf>
- Thomas, D. R. (2006). A general inductive approach for analyzing qualitative evaluation data. *American journal of evaluation*, 27(2), 237-246. <https://doi.org/10.1177/1098214005283748>
- Tipton, S. J., Forkey, S., & Choi, Y. B. (2016). Toward proper authentication methods in electronic medical record access compliant to HIPAA and CIA triangle. *Journal of medical systems*, 40(4), 1-8. <https://doi.org/10.1007/s10916-016-0465-x>

- Tjora, A. (2021). *Kvalitative forskningsmetoder i praksis*. (4. edition). Gyldendal.
- Trom, L., & Cronje, J. (2019, March). Analysis of data governance implications on big data. In *Future of Information and Communication Conference* (pp. 645-654). Springer, Cham.
https://doi.org/10.1007/978-3-030-12388-8_45
- Tse, D., Chow, C. K., Ly, T. P., Tong, C. Y., & Tam, K. W. (2018, August). The challenges of big data governance in healthcare. In *2018 17th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/12th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE)* (pp. 1632-1636). IEEE.
<https://doi.org/10.1109/TrustCom/BigDataSE.2018.00240>
- Ullah, F., Habib, M. A., Farhan, M., Khalid, S., Durrani, M. Y., & Jabbar, S. (2017). Semantic interoperability for big-data in heterogeneous IoT infrastructure for healthcare. *Sustainable cities and society*, 34, 90-96. <https://doi.org/10.3390/asi1040043>
- University of Bergen. (2021, July. 6). *Risk Matrix*. From:
<https://www.uib.no/en/hms-portalen/142418/risk-matrix>
- Winter, J. S., & Davidson, E. (2019). Big data governance of personal health information and challenges to contextual integrity. *The Information Society*, 35(1), 36-51.
<https://doi.org/10.1080/01972243.2018.1542648>
- Åm, H., Frøyhaug, M., & Tøndel, G. (2021). Helsedata som gullgruve? – forventninger til kommersialisering av helsedata i Norge. *Nytt Norsk Tidsskrift*, 38(1-02), 86-98.

APPENDIX A

Interview Guide

Interview guide

- Start with who we are
- Why we do this research
- How should the information that emerges in the interview be used (how many people will be interviewed, is this one for a public publication, the respondent is anonymous).
- If we find it interesting, follow-up questions will also be asked about the answers given
- Deviations may occur from the interview guide if we find it necessary.
- The interview format will be semi-structured formal interviews

Questions
Tell us a little about your background, education, and work experience
What is your current job description and what are your area(s) of responsibility?
From a health sector perspective; what opportunities does Big Data give us?
What challenges do you find connected to Big Data in the health sector? And what are your thoughts around the aspect of security of these data?
How important do you consider learning from earlier cyber security events, either attacks or attempts on attacks, in order to make today's systems more resilient to handle the current security threat picture?

Suggestions for follow-up questions

- Question:
 - Are there any areas in the value chain that you consider most vulnerable when it comes to the security of Big Data?
- Question:
 - From your perspective; what opportunities does Big Data give us?
- Question:
 - Do you have any examples of where the security in Big data has been too weak so that the application / system has had to be taken out of operation? If so, do you know what was the reason for this?
- Question:
 - How does the security change when processing sensitive personal data in the form of Big data?
- Question:
 - When it comes to Big Data, are there any challenges that stand out especially when it comes to processing sensitive personal data?
 - What kind of measures are being implemented to prevent these challenges? Examples?
- Question:
 - Are there any important laws that need to be considered when processing such information?
- Question:
 - What kind of cloud service would possibly be used for such a purpose?
- Question:
 - Do you have any concluding remarks or suggestions?