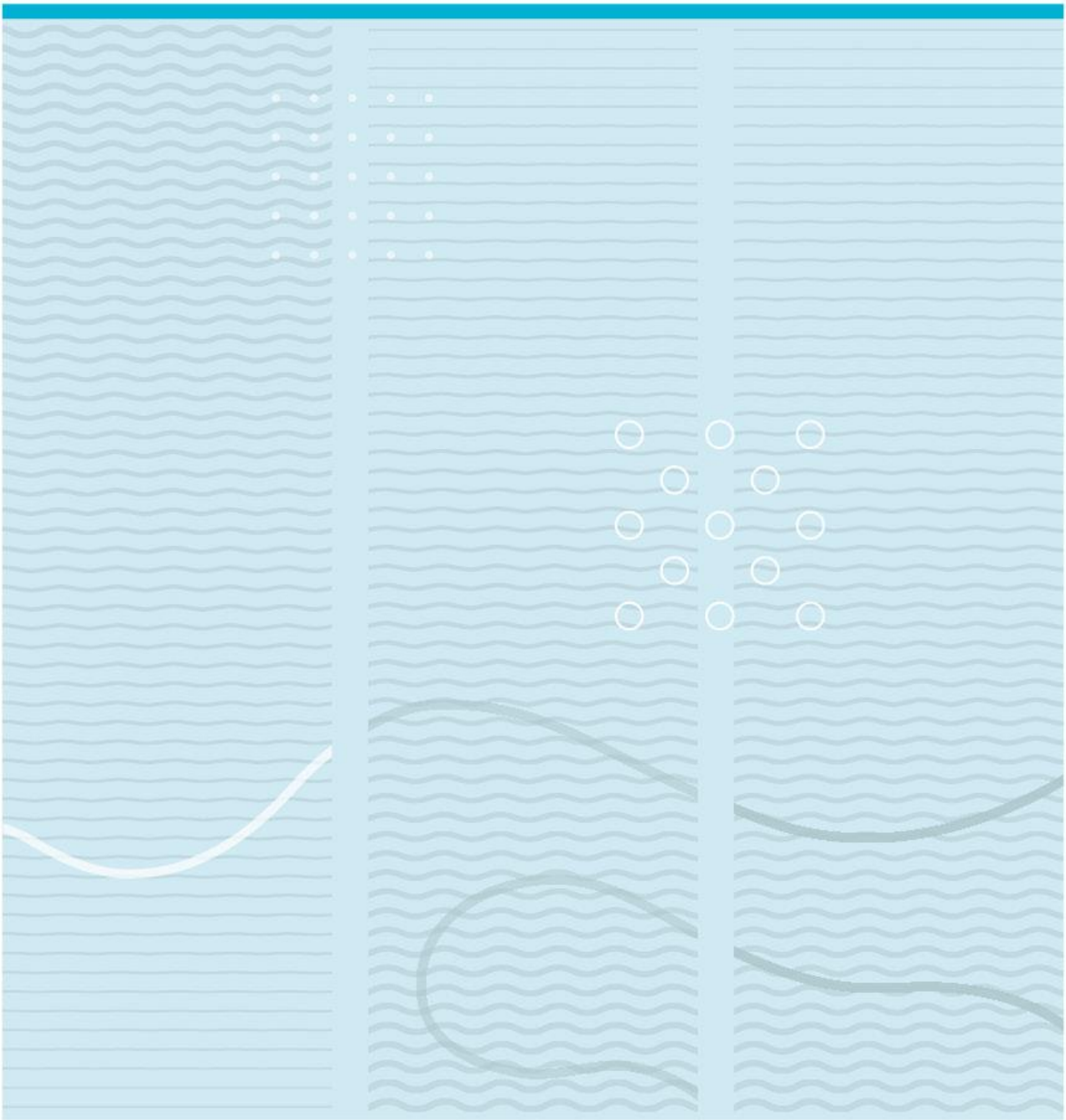
University of South-Eastern Norway
Faculty of Business, Marketing, and Law
–
Master’s Thesis
Study programme: Management Information Systems
Spring 2022

Gholamhossein Kazemi

Exploration of Health Data Management Systems From a Scandinavian Point of View



University of South-Eastern Norway
Faculty of Business, Marketing, and Law
Institute of Business
PO Box 235
NO-3603 Kongsberg, Norway

http://www.usn.no

This thesis is worth 30 study points

Abstract

This thesis explores blockchain-based health data management systems and their development factors in the context of smart city assets. Early designs, recent improvements, and innovative solutions for health data management will be explored from the point of view of scholarly publications, and patient-centricity will be explored from the point of view of the Twitter users from Norway, Sweden, Denmark, and Finland. In addition, the features and challenges of blockchain-based solutions will be examined based on the General Data Protection Regulation act and Regulations for the Directorate for e-Help of Norway. A thorough systematic literature review, detailed Natural Language Processing text and attitude mining, and Latent Semantic Space correlation examination has been conducted and a conceptual smart asset development framework for health data management has been proposed from a Scandinavian point of view. Moreover, this thesis proposes a conceptual patient-centered blockchain-based architecture for the development of current health data management in Scandinavia.

Contents

Acknowledgements

I deliver my most sincere gratefulness to my family and all the instructors that I have had during my entire education. The last two years of this journey have been a remarkable experience for me in which I have received nothing but support from both family and inspiring professors. I am highly grateful for the supervision of Doctor Karen Stendal who paved the way for my thesis from the start of my studies at the University of South-Eastern Norway, and for the support of Doctor Shegaw Anagaw Mengiste and Doctor Anh Nguyen Duc who helped me gain academic experience and hope. Last but not the least, I am thankful for the great atmosphere provided by all the members of the University of South-Eastern Norway.

Ringerike, Norway, 15th May 2022
Gholamhossein Kazemi

1 Introduction

Lately, a new normal has been defined for cities and their inhabitants, deep digitalization processes have taken place, and resilience measures have changed (IMD, 2021). In this regard, as a sample of Scandinavian countries, residents of Norway have perceived digital health services as one of the most prior and urgent indicators for their country (IMD, 2021). In the era of digitalization, healthcare has become highly dependent on data management. As a result, health data management systems have become increasingly important in cost reduction and treatment improvement of healthcare procedures (Mondal et al., 2019). This digital transformation of healthcare has empowered healthcare with diverse digital services like electronic health records, patient monitoring, family-rooted disease diagnosis, treatment enhancement, integrated health datasets, etc. However, challenges like big data, reliability and security have also risen simultaneously (Mondal et al., 2019). Plus, the International Institute for Management Development (IMD) reports that, as a consequence of digitalization, concerns have been raised regarding personal freedom limitations and potential misuse of the personal data collected during the digitalization process (IMD, 2021).

In addition, recently, data subject-centricity has been added to the criteria of any eligible health data management system as far as Korea, as a developed country, has changed its governmental attitude towards its data management systems in the medical, financial, public, logistic, cultural, communicational, educational, and energy field (Choi et al., 2021). This global movement toward data subject-centricity in healthcare has been named differently across the world, for instance, “MyData” is the South Korean context, and others are named “Self Data”, MiData”, “MesInfos”, “My Health Bank”, “Internet of Me”, “Personal Data Economy”, and “Personal Information Management Systems”. Nonetheless, all of them share the same content, which is the rightfulness of data subjects in making proactive decisions regarding sharing and transferring their data (Choi et al., 2021).

According to the most recent patient-centric approach in the health sector, especially the health data management sector, it is believed that blockchain technology is capable of enhancing the security and reliability of patients’ data. The combination of blockchain and the transformation of healthcare to a patient-centered approach helps consolidate and exchange patient data across interoperable health systems (Hölbl et al., 2018). Briefly, blockchain is a decentralized mechanism that stores peer-to-peer transactions either in a public or private mode and can be utilized with different consensuses and smart contracts to better handle different types of data in a secure, efficient, and transparent way (Al-Asmari et al., 2021). However, while thinking of blockchain as a solution for health data management, it is important to notify challenges like cost, data size, and privacy (Alamri et al., 2021).

Also, another 2020 study discussing the benefits and challenges of blockchain applications in the health sector indicates that the issues of traditionalism of the current systems and immaturity of blockchain solutions may lead to problems such as lack of interest in self-managing health data, interoperability of complex health systems, cyber-attacks, and resource consumption (El-Gazzar & Stendal, 2020).

In addition, the design and development of blockchain-based healthcare applications without standards and regulations will impact the implementation (Khatri et al., 2021). In Norway, as an example of a Scandinavian country, a GDPR-compliant blockchain solution is required that is aligned with the regulations of the Directorate for e-health. For instance, the OmniPHR architecture provides a patient-centered interoperable exchange of health information but is not GDPR-compliant (Alamri et al., 2021). The European Parliamentary Research Service reports that the role of the data controller and the right of erasure defined by the GDPR act are two problematic requirements of any blockchain-based health information system (Hasselgren et al., 2020).

Finally, this study aims to explore health data management, blockchain, and regulatory matters from the point of view of the scholars and users in Scandinavia with the help of text and social media mining to extract the common beliefs around blockchain-based health data management systems and user-centricity. Furthermore, considering the significant role of a patient-centered health data management system in a Scandinavian context and taking into account the rich infrastructure and technological approach toward implementation of innovative solutions (IMD, 2021; Laurini, 2020), this study investigates the possibilities of conceptualizing a blockchain-based data subject-centered health data management system that is compliant with the GDPR and Directorate for eHealth regulations. Accordingly, the research question is:

What is the common belief around blockchain-based data subject-centered health data management systems, and what blockchain capabilities can resolve related challenges?

To find answers to the above research question, this thesis has used the smart asset development framework to explore possible blockchain-based developmental solutions for current health data management systems. The common scientific belief regarding such a solution will be extracted from prominent scholarly publications and the common user-centered belief will be extracted from Twitter users in Scandinavia. A complementary blockchain-based patient-centered literature investigation will be conducted based on the requirements of the GDPR and the Regulations of the Directorate for e-Health to propose possible conceptual development models from a Scandinavian point of view.

Accordingly, the rest of this thesis is organized as, respectively, literature review, methodology, results, discussion, implications and future work, and conclusion. In the literature review section, the review process and scholarly scientific materials will be explored. The methodology section reveals the thesis's approach and the process of data collection and analysis, and the result section presents the findings. Then the findings from the previous sections will be discussed and explored from a Scandinavian point of view in the discussion section, and a summary will be presented in the conclusion. Ultimately, the last section includes the limitations of the study and possible future paths.

2 Literature Review

Considering the topic of interest of this thesis, a thorough exploration of scholarly articles regarding the core concepts of data management and blockchain was required along with an emphasis on the Nordic approach toward health e-services delivery. As a result, guidelines introduced by Webster & Watson (2002) were used to conduct a systematic literature review covering all the grounding designs of data management in the framework of smart asset development. In the rest of this section, grounding frameworks of data management, their developments, and capabilities and challenges of blockchain in revolutionizing data management will be scrutinized based on the laws and regulations of health data management in Norway as a Scandinavian sample country.

2.1 Health Data Management System: A Smart Asset

Smart City Index (SCI) report takes into account the perception of those who live and work in the cities and defines a smart city as "*an urban setting that applies technology to enhance the benefits and diminish the shortcomings of urbanization for its citizens*" (IMD, 2021, p 3). These indexes assess the perception of residents on factors related to structures and technology applications implemented in their city. The former relates to the existing infrastructure and the latter refers to the technological provisions and available services for the residents. Each of these main factors includes five key areas of health and safety, mobility, activities, opportunities, and governance (IMD, 2021). This report also indicates that Norway is rated "AAA" for the structures and "A" for the technologies. Medical services provision satisfies 78.8% of the respondents and 69.6% state that online medical arrangements have improved access. Moreover, an average of 65.7% of the respondents indicate that they are willing to concede personal data to reduce traffic, are comfortable with face recognition to reduce crime, and feel that information availability has increased their trust in the authorities (IMD, 2021).

Nordic countries are in an increasingly ongoing collaboration with information technology companies for the enhancement of data analytic products in favor of cost-saving and cost-efficiency in services, and health services are one of five public services that work in the framework of smart cities (Choroszewicz & Alastalo, 2021; Gil-Garcia et al., 2016). One of these data analytic products or smart assets is data management systems as they define it as "*a data analytics product that provides management with information about clinical and financial aspects of organizational management*" (Choroszewicz & Alastalo, 2021, p. 1). As a result, a holistic framework is required to understand the drivers of smart assets. As illustrated in Figure 1, policy, technology, and community are the drivers of any smart city asset's development, and health data management systems are one of these assets (Yigitcanlar et al., 2018).

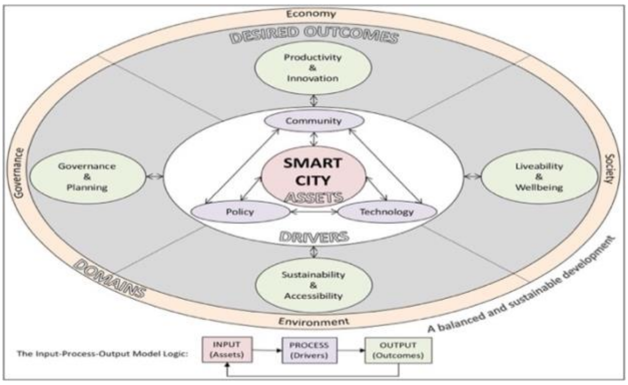


Figure 1 - Smart City Development Framework (Yigitcanlar et al., 2018)

Moreover, integration, innovation, evidence-based decision making, citizen centricity, sustainability, creativity, effectiveness, efficiency, equality, entrepreneurialism, citizen engagement, openness, resiliency, and technology savviness are the dimensions of smartness in governments (Gil-Garcia, as cited in Ismagilova et al., 2019). However, it is almost impossible to find all these dimensions in every asset. Assets include the dimensions that help their strategizing, evaluation, and development the most (Gil-Garcia et al., 2016). Consequently, any health data management system as a smart asset that delivers digital public health services must consider the drivers and related dimensions of smartness in their designs. In a health data management system context, policies refer to GDPR and Directorate of eHealth regulations, technology refers to the infrastructure, and community refers to the users of health data management systems. In the following, health data management systems are being explored in detail to find out how to relate their elements to these development factors and smartness dimensions.

Data is the core of any data management system and is a symbolic representation of observable or non-observable properties. In other words, data are the givens of any kind that leads to information, knowledge, and wisdom, as illustrated in Figure 2 (Fricke, 2009). Data management systems store, process, retrieve, and deliver structured, unstructured, semi-structured, and streaming data to support data organization (Modal et al., 2019). Every data management system includes three main processes, namely storing, processing, and accessing. The storing process can adopt different technologies based on the defined procedures of data storage. Also, data processing minimizes the data service costs by handling the volume and reduction of unwanted data, and data accessing is focusing on the retrieval of data from the data warehouse (Mondal et al., 2019). Whereas, in an everchanging environment with the continuous evolution of data in vastly different areas, dynamicity becomes crucial. Dynamic Data Management Systems (DDMS) try to overcome this challenge by providing data management based on four criteria. DDMSs take into account that stored data is large and changing all the time, views' maintenance dominates ad-hoc querying, data accession privileges views and simple processes over heavy updates, and updates happen via an update stream (Kennedy et al., 2011).

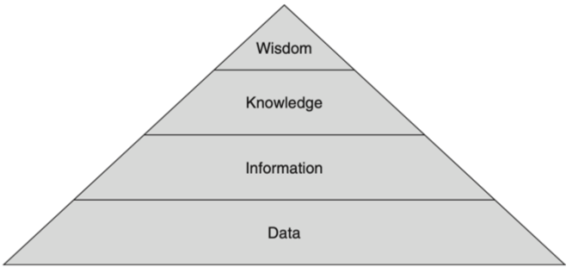


Figure 2 - The DIKW Pyramid by Ackoff in 1989- (Fricke, 2009)

2.2 Health Data Management System: Early Designs

Data management systems (DMSs) investigate patterns and trends that help data minimalization and outcome improvement by emphasizing business value creation out of social services data. DMSs contain online dashboards that allow users to track data flow, decisions, paths, outcomes, and costs in real-time (Choroszewicz & Alastalo, 2021). In other words, DMSs work like boundary objects that unify diverse interests of stakeholders in a social world with shared interests. These interests shape the structure of DMSs (Choroszewicz & Alastalo, 2021). Normally, a DMS consists of storing, processing, and accessing elements that operate based on five main principles, namely data process policy, data classification, data ownership, data curation and archiving, and data access and sharing (Hu et al., 2021; Mondal et al., 2019).

One of the very first mentions of a clinical data management system refers back to the 1960s when the need for standardization, quality control, and retrievability of clinical data emerged (Greenes et al., 1969). Terminology was not standard, data was not organized and well-formatted, and notes were sometimes unreadable (Greenes et al., 1969). The early designs of health data management systems were used to take into account user interface requirements, variable data handling, hierarchical database access, multi-user access, large storage capacity, low processing power, and high-level language (Greenes et al., 1969). However, in the next decades, more factors were added to the center of attention of such system designers. Factors like high-quality data acquisition, commercial uses of clinical data in pharmaceutical research, global and platform-independent data standards, and interoperability of healthcare (Fu et al., 2010). Some of the recent cloud health-applicable data management systems are, namely Amazon, Hadoop, and Microsoft Health Vault. For example, Amazon S3 is used to find the cheapest e-health services in developing countries, Hadoop is used for real-time health data monitoring, and Picture Archiving and Communication System (PCAS) is used for storing and sharing medical images (Mondal et al., 2019). However, outsourcing makes patients' sensitive data vulnerable to security, privacy, and integrity failures (Ismail et al., 2020).

Personal health data is considered the most crucial element of any health data management system and requires mutual trust between the patient and other stakeholders for secure sharing. In general, health data management systems include two stakeholders, namely the data owner and data requester while the former refers to the patient and the latter refers to anyone else who is working on the other side of the system (Ismail et al., 2020). The process of data sharing in HUMARIS, one of the very early models of health data management systems, was divided into four phases of data acquisition, request for materials, shipment, and feedback and control (Minckler et al., 1967). Clinical data acquisition-request was expected from diverse sources like the patients, doctors, consultants, laboratories, etc.; hence, the need for a universal clinical data structure, user-friendly interface, and miscellaneous kinds of data export emerged so fast (Greenes et al., 1969). Simultaneously, the patient's data sensitivity and volume brought up the urgency of a large multi-level access database of medical records. In addition, vital and emergency illness cases required a system capable of handling a large volume of computing in the shortest possible time (Greenes et al., 1969).

The next models optimized the phases into acquisition, storage, and retrieval with a greater focus on standardization, quality control, and retrievability. The more clinical data management was developing, the higher the pressure on high-quality data was becoming from the governmental and medical research organizations (Fu et al., 2010). Over time, countries like the United States of America and China started huge investments in clinical data management systems and even in most cases stopped receiving paper-based data submission (Fu et al., 2010). Consequently, global consortiums were being established to help the standardization and interoperability of such systems. In this respect, Clinical Data Interchange Standard Consortium (CDISC) introduced standards for the exchange, storage, and archival of health data. Standards like platform-independent Operational Data Model (ODM) for data exchange and storage, Clinical Data Acquisition Standards Harmonization (CDASH) which recommends the minimal data collection in 16 fields such as demographic, therapeutic, safety, etc., or Standard Data Tabulation Model (SDTM) concerning standard data tabulation structures (Fu et al., 2010). These new movements in the development of health data management systems were accompanied by a revolutionized system architecture for such systems, as illustrated in Figure 3.

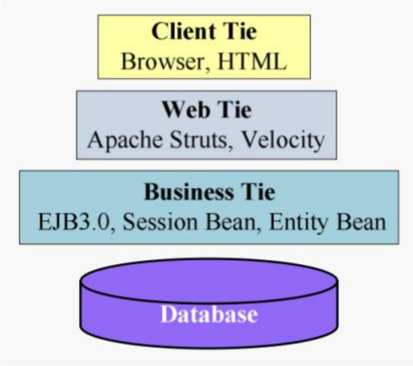


Figure 3 - Clinical Data Management System Architecture (Fu et al., 2010)

This four-layered architecture consists of layers, namely client tie, web tie, business tie, and database tie. This is a web-based and platform-independent architecture with a relational back-end database that lets every layer run on different hardware platforms (Fu et al., 2010). This system has 4 modules of study management, data capture, data management, and security. The study management module is used for the creation, modification, and control of different medical research studies, the data capture module is the most significant one which is in control of data collection from patients, the data management module controls data export, sampling and report, and the security module manages roles, log-in, access layers, etc. (Fu et al., 2010). Nevertheless, early designs had some flaws. For instance, they were vulnerable to counterfeits, patients had no full control over their data, data were sold most of the time to third-party companies, and there was only one central authority that could lead to a single point of vulnerability (Aggarwal & Kumar, 2020).

2.3 Health Data Management System: Recent Improvements

Needless to state, the health data management system's development is highly correlated with the advancements in technology and their adoptions' alignment with biomedical care and research

(Ismail, et al., 2020). Figure 4 shows the transition of health data management systems throughout history. Technological advancements have pushed health data management systems from paper-based to computer-based, and then cloud-based, IoT-based, and blockchain-based systems which are more accurate, efficient, diagnostic, error-less, and cost-effective (Ismail et al., 2020). As a result, over time, two categories of e-health and m-health emerged including the concepts of the web, patients, professionals, social networking, health information content, collaboration, coordination, integration, and interoperability; however, concerns regarding privacy, security, public insight, and patients' participation in accessing and monitoring were rising concurrently (Ismail, et al., 2020). Furthermore, over these years, the definitions of health data management systems have also transformed from merely medical practice and learning-based to more patient-centered and research-based (Ismail et al., 2020).

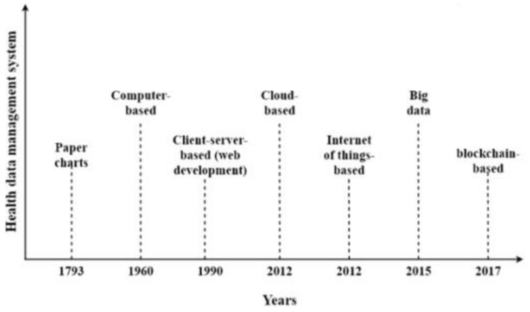


Figure 4 - Transitions of Health Data Management Systems (Ismail et al., 2020)

Each of these phases of transition had flaws. For instance, in the first web-based systems, patients were not able to trace their data, and security and privacy were at the risk of a single point of failure. Moreover, there was not a cohesive holistic view of patients' data. Also, could-based systems were vulnerable due to the possibilities of loss of data control and governance, and the necessity of steady internet connection and reliable data (Ismail et al., 2020). In addition, early IoT-based systems were not secure enough and caused privacy problems for patients. Likewise, big-data-based systems were struggling with time consumption for data preparation for processing, and analysis. They were complex and expensive, and maintaining security and privacy during the extraction of useful data was challenging. Finally, blockchain-based solutions appeared to be energy-consuming and time-consuming (Ismail et al., 2020).

Taking into account the conceptual and practical transitions of health data management systems, a seven-element requirement framework has been introduced that includes medical data records, real-time data, patient participation, sharing, security, privacy, and public insight (Ismail et al., 2020). Respectively, medical data record reflects the identity and health status of a patient through personal and demographic data, historical and ongoing medical data, and diagnostic tests and results. Real-time data emphasizes accession to updated data for the enhancement of health care, and patient participation allows patients to track down the procedures of processing their data. A user-friendly system elevates both engagement and trust of patients by letting them access, monitor, and control their data, which results in better diagnostic and treatment services (Ismail et al., 2020).

Plus, sharing enables cross-sector data accession in a classified manner; however, it requires interoperability through the acquisition of standard formats of storage, management, and sharing of medical data. Sharing is vital for treatments dependent on cross-sectional cooperation and coordination. In 2018, 503 security breaches impacted over 15 million patients around the world (Ismail et al., 2020). Accordingly, due to the sensitivity of the personal and medical data of an individual, security plays a very crucial role in health data management systems. Patients' identical and medical privacy is also a major requirement to avoid fraudulent and fake medications. Finally, public insight, which is a result of big health data analytics, facilitates predictive diagnosis and evaluation of health conditions through the extraction of correlations and meaningful patterns from big health data. Although public insight helps agile improvement and timely response of health care, it is highly dependent on personal medical data and requires a well-defined balance between privacy and transparency (Ismail et al., 2020).

As mentioned above, big data plays an important role in health data management systems. It is related to public insight, security, privacy, and transparency (Ismail et al., 2020). Big health data management systems have four entities which are data source, data manager, data warehouse, and service consumer (Mondal et al., 2019). Data source refers to any incoming data in the form of text, video, image, audio, etc., and the data manager is responsible for the collection, integration, preprocessing, and storing of the data in the data warehouse. Ultimately, service consumer requests data based on their needs and retrieve it from the warehouse for further analysis, control, services, etc. (Mondal et al, 2019). Nonetheless, big data has its challenges like storage capacity, the ratio of useful/useless data, data concurrency and consistency, and data type diversity (Mondal et al., 2019).

2.4 Health Data Management System: Patient-Centricity, Security, and Privacy

Ackoff's hierarchy explains the functional transformation of data to information, information to knowledge, and knowledge to wisdom step by step and at the same time conveys the fact that each lower part of the hierarchy exists in the upper part genuinely (Fricke, 2009). Accordingly, in a smart context, on one hand, data is considered an element of the knowledge creation process, and knowledge management and creative solutions are necessary to sustain smartness (Fricke, 2009; Laurini, 2020). On the other hand, citizen-centricity and technology savviness dimensions impose the burden of a citizen-centered approach on the shoulders of the government in implementing creative technological solutions (Laurini, 2020).

As a result, health data management systems as the technological handlers of personal health data and their implementation approach are critical in a Scandinavian smart context. Considering the high-tech infrastructure richness and security significance of personal data in a Scandinavian smart context, a secure citizen-centered health data management system plays an important role in delivering electronic health services (IMD, 2021).

Health data management has experienced a paradigm shift regarding its active-role centricity since its initial implementations. In the 1970s, such systems were organization-centered; however, in the mid-2000s the concept of user-centricity emerged (Pramanik et al., 2017). Different worldwide contemporary governmental movements toward the protection of individuals' rights regarding their data try to defend the very same right of giving the data subjects control over their data in a transparent and decentralized ecosystem (Choi et al., 2021). For instance, in the United Kingdoms of England people can voluntarily share their health data with health service providers for long-term care and telemedicine, or in Sweden, patients can choose what entities of their health data should be shared with health service providers, or Australians can choose which medical institutions can access their health data (Choi et al., 2021). Utilization of personal health data has a direct relationship with the willingness or resistance to disclosure of personal medical data. This utilization must consider consent-based use of personal data as a vital fundamental element. Such utilization requires a specific personalized service architecture for patient-centered health data management (Choi et al., 2021).

The process of consent management can be divided into three steps. First, patients, as the data subjects, download or collect their health data on a platform, then they agree to the terms of data sharing on such a data management platform, and finally, they define their customized consent to data sharing and accession for third-party stakeholders (Choi et al., 2021). Third-party stakeholders may vary from diagnostic institutions or lifestyle management agencies to healthy food consultancy agencies and medication institutions. Accordingly, the core functions of a patient-centered health data management system are, namely consent management, data collection from sources, and data sharing with third parties (Choi et al., 2021). The patient-centered approach is necessary for a digital democratized facilitation of clinical trials, frontline care, data surveillance, medical billing, telemedicine, drug delivery, treatment, and strategizing (Jabarulla & Lee, 2021). Plus, since the patients are the owners of their health data, a patient-centered approach democratizes health data monetization and protects patients' privacy by building trust through a transparent procedure of data storing and sharing (Jabarulla & Lee, 2021).

An experimental implementation of a patient-centered health data management system in South Korea revealed that based on socio-demographic results women tend to participate more in health-related data management systems, people in their 30s have the highest rate of participation., and people over 50 have the lowest rate of participation (Choi et al., 2021). Moreover, the deeper the depth of accession permission, the lower the rate of consensus grant. Around 94% granted data search permission, 91% granted download permission, 84% granted sharing permission, and 79% granted commercial permission (Choi et al., 2021). Although these rates of access permission may vary in different categories of health data like cancer, genetics, etc., the overall attitude toward the use of a patient-centered health data management system is positive (Choi et al., 2021). Also, a recent pandemic study revealed that a patient-centered approach is crucial in critical situations that require individual track of treatment (Jabarulla & Lee, 2021).

A 2015 United States report from the Office of the National Coordinator for Health Information Technology (ONC) indicated that information blocking codification placed by the ICT vendors and providers limits patients' access. This creates disconnected information islands and decreases patients' engagement (Hylock & Zeng, 2019). A patient-centered health data management system elevates personalized treatments by putting the patients at the center of attention and in charge of the control, access, and sharing of their data; however, a shift of approach is required from the current centralized to decentralized health care system (Jabarulla & Lee, 2021).

The US Federal Information Security Management Act defines the security objectives of personal health data as, respectively, confidentiality, integrity, and availability; while health data is supposed to be the most vulnerable, easy, and tempting type of data for criminals, as a 2015 study indicates that 43% of data breaches had been recorded in the health sector (Shakil et al., 2017; Wu et al., 2016). Health data annual growth will reach the rate of 36% by the end of 2025 which brings up struggles with latency, scalability, and management (Chkirbene et al., 2020). Health data is easier to access than financial data, easier to breach due to lack of security measures, and slower to theft detection. As a result, it tempts the data criminals the most (Shakil et al., 2017). Moreover, breached data can be used for many other malicious acts like insurance fraud, illegal drug fake prescribing, credit card fraud, etc. As a result, unstructured and unencrypted growing big health data require security at the most possible level. In this regard, from an authentication point of view, unauthorized access to health data should be prohibited, and biometric signatures are a possible solution to this phenomenon due to their social acceptability (Shakil et al., 2017). Big health data is also intertwined with the concept of the Internet of Things (IoT) or real-time data collection and access. Thereby, it is crucial to consider authorized and secure access not only from a user point of view but also from an IoT device point of view (Meng et al., 2019).

Traditional security and privacy measures implemented in the current health data management systems have shown flaws that require edge-of-art alternative solutions to be tackled (Rajawat et al., 2022). Most of the existing health data management systems use centralized servers that are prone to single point of failure vulnerabilities, insider attacks, and loss of control over outsourced data while decentralized solutions can tackle such a flaw (Shi et al., 2020). AI, big data, Blockchain, distributed ledgers, smart contracts, consensuses, etc. help overcome the challenges of patient-centricity, security, and privacy; however, they bring up other challenges as well which will be explained in the following. Before diving into the blockchain-based solutions for health data management systems, regulations of health data management from the point of view of the GDPR act and the Directorate for e-health of Norway will be explored in the next section.

2.5 Regulations

As mentioned before, the core element of any health data management system is data and personal health data in Norway, as a sample of Scandinavia, is protected by the GDPR act and regulations of the Directorate for e-health (EU, 2022; Fricke, 2009; Vigot & Bussche, 2017). Nordic countries are well-known international leaders of public service digitalization with the help of combinable data they collect from citizens. Collected data elevate citizens' engagement with public institutions like healthcare (Choroszewicz & Alastalo, 2021). Accordingly, the GDPR act and regulations of the Directorate for e-Health are explained in the following.

2.5.1 General Data Protection Regulation Act (GDPR)

As the most important pillar of the data management systems, personal data is protected by the General Data Protection Regulation (GDPR) act. GDPR consists of 11 chapters, 99 articles, and 173 acts (EU, 2022; IMD, 2021). Although Norway is not a member of the European Union (EU), GDPR is effective in Norway, Sweden, Denmark, and Finland, and all organizations must be compliant (EU, 2022). Traditional data management systems were not supposed to be GDPR compliant; however, with GDPR penalizing personal data violations up to 20 million Euros or 4% of the global annual turnover, it is necessary to design and implement a GDPR compliant and secure system architecture for processing personal data (Nam et al., 2020). Especially when it comes to a data management system that processes personal data on behalf of governmental organizations for providing digital public goods like digital health services.

GDPR became effective in 1995 as a replacement for the EU's old-fashioned data protection directive; however, then in the digital age, on 25th May 2018, a newly updated GDPR act was defined for easy and cheap data protection compliance in all the sectors of industry, including healthcare (Haque et al., 2021). GDPR clarifies the patient's rights to store, access, and share their data, and considers the data to be real-time and portable (Alamri et al., 2021). GDPR act defines the stakeholders of any data processing procedure as the data subject, data controller, and data processor, and specifies personal data as "*any information that relates to an individual who can be directly or indirectly identified*" (EU, 2022; Vigot & Bussche, 2017, p 11).

GDPR explicitly defines the data subject as the owner of the information, the data controller as "*a natural or legal person, public authority, agency or other body that alone or jointly with others determines the purpose and means of data processing*" (Vigot & Bussche, 2017, p 17), and the data processor as "*a natural or legal person, public authority, agency or other body that processes personal data on behalf of the controller*". Accordingly, the data controller can either process data internally or outsource it to an external data processor (Vigot & Bussche, 2017, p 20).

The data protection principles introduced by GDPR emphasize lawfulness, fairness, and transparency of data processing, purpose limitation, data minimization, accuracy, storage limitation, integrity and confidentiality, and accountability (EU, 2022; Vigot & Bussche, 2017). Briefly, processing must be lawful, fair, and transparent to the data subject, and the purpose of the process must be legitimate and clear to the data subject. Moreover, data should be collected only as much as absolutely necessary for the specified purpose and should be kept accurate and up to date. Collected data should be stored only as long as necessary for the specified purpose, and processing must ensure security, integrity, and confidentiality. Eventually, the data controller must be able to demonstrate GDPR compliance with all the aforementioned principles (EU, 2022; Vigot & Bussche, 2017). Article 17(2) of GDPR entitles the data subjects with the right to be forgotten, even in the cases in which the data has been published publicly. In this respect, the data controller has to inform all the third-party data processors about the request to erase data and be forgotten (Vigot & Bussche, 2017).

Health data is defined as "*personal data related to the physical or mental health of a natural person, including the provision of healthcare services, which reveal information about his or her health status*" in Article 4(15) of the GDPR act (Hasselgren et al., 2020, p1). Any health data subject has the right to access, delete, or move their health data, and the right to be informed if their data is being collected by anyone else than themselves (Alamri et al., 2021). For example, if an organization intends to sell or move a patient's data to another organization, it must clearly state the procedures of the data sharing process and act according to the Data Processing Agreement (DPA) including processing instructions, confidentiality, security, privacy, and most importantly consents (Nam et al., 2020). According to Articles 4(11) and 7(1) of GDPR, consent is any freely given, specific, informed and unambiguous indication of the data subject's wishes by which it, by a statement or by a clear affirmative action, signifies agreement to the processing of its personal data, and the controller is the responsible body for the proof of consent-based processing of data (Vigot & Bussche, 2017). In case of any violation related to consent, privacy, or data breaches, Article 82 entitles the data subject to the right to claim compensation for material or non-material related damages (Vigot & Bussche, 2017).

Moreover, patients should be able to update their data and limit third-party accession to it in favor of their interests. Also, they have the right to make decisions regarding commercializing their data and should be informed in less than 72 hours if their data is breached (Alamri et al., 2021). In a healthcare context, portability of data refers to the features of interoperability and user-centricity of health data management systems, which are musts based on the GDPR acts. For instance, such systems can use the HL7/FHIR standard that ensures semantic data interoperability and exchange (Alamri et al., 2021).

A GDPR-compliant data sharing process includes three steps. First is the data-sharing request in which a user requests to retrieve data from the server based on the DPA; second is the data-sharing approval in which the server governs and validates or invalidates the request and its scope compliance with the DPA; the third is the data-use in which the server provides the access tokens to the data (Nam et al., 2020). Otherwise, the data subjects can bring their complaints to a supervisory authority in case of suspicious or violent processes from an organization claiming to be GDPR-compliant (Esmel et al., 2021). In addition, the service provider, which is the health sector in this context, is the responsible body for data collection, control, and process. In this regard, the data collector is responsible for data movement throughout and across interoperable systems, and the data processor oversees data manipulation, sharing, and deletion under the supervision of the data controller. However, it requires the data controller to be aware of the full procedures of the data process (Haque et al., 2021).

2.5.2 Directorate for e-Health Regulations

Other than the GDPR act, which is effective in European Union, each country has a responsible body for handling digital health services and its regulations. Here the scope of the study is Scandinavia, but the main focus is on Norway as a sample of Scandinavia. In this regard, the responsible organization for e-health regulations in Norway is the Directorate for eHealth which defines the goal architecture for data sharing in the healthcare sector. Directorate for eHealth in Norway in the Patient Records Act 18, Patient and User Rights Act 5-1, Privacy Ordinance Article 15, and Health Personnel Act 41 clearly states that "*citizens have a statutory right of access to information that is registered about them in treatment-oriented health registers*" (Directorate for eHealth, 2021, p 90). Thereby, patient-centered health data management systems are necessary to be implemented in Norway as a Scandinavian country to comply with the GDPR and Directorate for eHealth regulations.

The Directorate for e-health published a goal architecture for data sharing in March 2021 focusing on the security, privacy, and confidentiality aspects of sharing health data (Directorate for eHealth, 2021). This architecture consists of two main actors, namely digital health service providers and citizens. Cross-sectional data sharing is anticipated in advance by suggesting using common data-sharing components. The goal architecture is a Norwegian translation of the European Interoperability Framework (EIF) that has national, regional, and data controller levels; however, it is flexible to meet the needs of both large and small actors. Moreover, any change at a national level affects all the lower levels also (Directorate for eHealth, 2021).

Common components of the architecture mentioned above are, respectively, 1-HealthID for identifying health workers, 2-Citizen STS for identifying patients, 3-Joint Application Programming Interface (API) services for national e-health service providers, 4-Common authorization, consent handling, logging, etc., 5- Common health information-seeking component, and 6- Common API catalog (Directorate for eHealth, 2021). In addition, patients are the most important part of the architecture who are entitled to the right to follow their treatments, decide about their treatments, repeat advice during the consultation, seek reassessment, choose selective accession for anyone other than themselves, and know the holders of their information (Directorate for eHealth, 2021). Six possible intentions of health data sharing in this architecture are listed as 1- Self-service access for sharing information with health or healthcare evaluator companies, 2- Provisional cross-sectional health data sharing for updating health information, 3- Sharing health data with citizens through applications and websites, 4- Sharing administrative basic health data, 5- Sharing health information with the public agencies such as Police, Tax agency, etc., and 6- Sharing health information for research and quality improvement (Directorate for eHealth, 2021).

Noteworthy to state, any kind of data sharing must be conducted in a mutual trust manner that has been made through a common trust establishment component. Any company intending to share or receive information must establish trust with the other side in advance (Directorate for eHealth, 2021). This trusted sharing can take place only and only with the consent of the patient. There is only one exemption for unconsented data sharing, which is the situation in which the life of the patient is in fatal danger (Directorate for eHealth, 2021).

2.6 Blockchain-Based Data Management System: An Emerging Asset

Blockchain has vastly different applications in healthcare such as in prescribing, supply management, billing, contracting, clinical trials, health insurance, audits, data management, etc. However, data management is the most important one considering its sensitivity (Hölbl et al., 2018). Blockchain provides various patient-centered data access, control, and sharing methods with robust stability against attacks and failures (Hölbl et al., 2018). A usable blockchain-based system empowers the users with unlosable data ownership and incentivized data sharing (Shrestha & Vassileva, 2019). To identify the cons and pros of using blockchain in health data management, in the following, the blockchain concept and its health data management applications will be explored in detail.

2.6.1 Blockchain Concepts

The US National Institute of Standards and Technology (NIST) describes blockchain as a tamper-resistant, distributed, decentralized, and consensus-based technology capable of recording and storing transactions in a shared-community manner (Yaga et al., 2018). Although its emergence refers back to the late 1980s, it flourished in 2009 with its cryptographic financial application. Bitcoin was the cryptocurrency that brought blockchain to the light (Yaga et al., 2018). Blockchain technology has gone through three revolution phases. Blockchain 1.0 is all about Bitcoin and cryptocurrencies, and blockchain 2.0 is tied with registering, confirming, and transferring contracts and properties with the help of smart contracts. Finally, Blockchain 3.0 extended the applications of blockchain and moved it from an only financial technology to a governmental, health-related, scientific, educational, etc. one (Gatteschi et al., 2018). Transparency and security are the prominent features of blockchain that are enabled by ledger, cryptography, shared consensus, and distributed characteristics. Ledger provides a full transaction history without overwriting, cryptography ensures security and attestability of data, shared consensus ensures transparency and trust, and distributivity elevates attack resistance (Yaga et al., 2018).

Generally, blockchains are divided into permissionless or public, permissioned or private, and consortium or federated blockchains (Dinh et al., 2018). Permissionless blockchains are open-source software that let anyone on the network write on or read from the ledger and issue transactions; however, consensuses like Proof of Work (PoW), Proof of Stake (PoS), etc. are placed to prevent malicious transactions. (Yaga et al., 2018). On the other hand, permissioned blockchains are either open or closed source software that can restrict reading and writing access based on authorization. Moreover, consensuses are not mandatory since malicious transactions revoke the authorization by shedding light on misbehaving. Consequently, permissioned blockchains are faster and require less computational power (Yaga et al., 2018). Finally, consortium blockchains are the same as permissioned blockchains but with more than one authority each governing the blockchains' propagation both locally and in cooperation (Dinh et al., 2018; Yaga et al., 2018). Blockchains are working in six layers which are, respectively, data, network, consensus, contract, service, and application layers (Yang, 2019). The data layer includes the main chain structure and blocks, the network layer includes P2P relations and verification mechanisms, the consensus layer defines the node protocols, the contract layer includes incentives, the service layer includes the provided product or service, and the application layer indicates the context of the service (Yang, 2019).

Each blockchain consists of cryptographic hash functions, transactions, asymmetric-key cryptography, address, ledgers, and blocks. In this regard, the cryptographic hashing function applies encryption and decryption to data (Yaga et al., 2018). Transactions are the interactions between the parties. Asymmetric-key cryptography uses a mathematically related public and private key for the encryption and decryption of data to build trust between the users who do not know each other. This way the transactions stay public and transparent; however, only the ones with the private key can sign the transaction. (Yaga et al., 2018). Addresses work differently based on the implementation. They can show the start-end point of transactions or storing point, etc. Ledgers are a collection of transactions, and blocks contain block header and block data. Transactions are added to the blocks when a publishing node publishes a block (Dinh et al., 2018; Yaga et al., 2018). The Block header consists of content metadata and the block data is a list of validated and authenticated transactions. Metadata includes block number, previous block header's hash, current block's hash, timestamp, size, and maybe a nonce value. Block's data also includes a list of within-block transactions and ledger events, and maybe other data (Yaga et al., 2018). Figure 5 shows a chain of blocks.

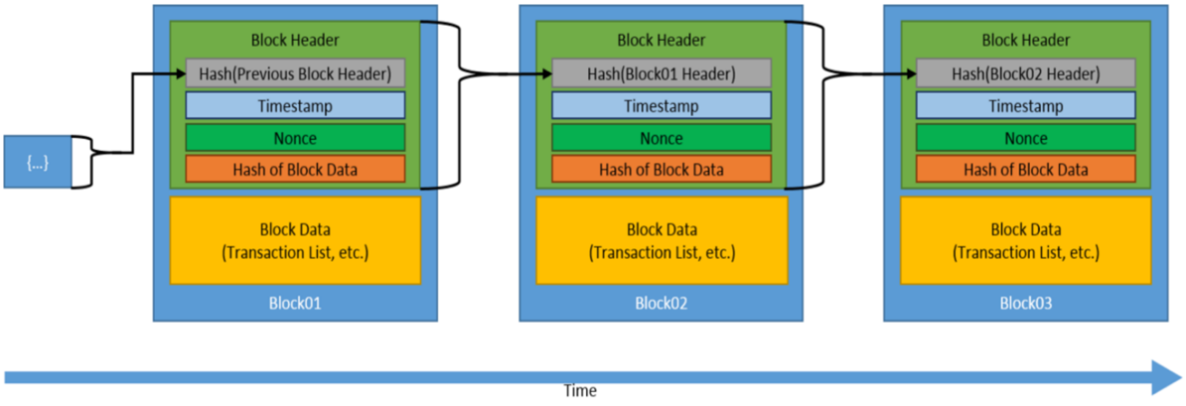


Figure 5 - Generic Chain of Blocks (Yaga et al., 2018)

Some consensus algorithms are being explained to get to know more about how blocks get propagated. The key factor of block propagation is the consensus algorithm (Yaga et al., 2018). Consensuses help agreement on the initial stage of the system, block propagation, block linkage, and independent block verification. For example, the Proof of Work (PoW) consensus is based on solving puzzles (work). This way publishing nodes can identify valid blocks that have solved the puzzle. A great example of PoW-based blockchains is the Bitcoin cryptocurrency (Dinh et al., 2018). The Proof of Stake (PoS) consensus validates the users and blocks based on the amount of the stake they have put in the system. Stakes can be cryptocurrencies but are not spendable (Yaga et al., 2018). The Round Robin (RR) consensus works better in permissioned blockchains and propagates blocks in turn. It prohibits a node to publish the majority of the blocks. The Proof of Authority (PoA) or Proof of Identity (PoI) consensus works based on real-world identification. If a node wants to publish a block, it should prove its identity by staking its identity or reputation (Dinh et al., 2018).

The Delegated Proof of Stake (DPoS) consensus reduces the computational power by using the shareholding nodes' votes to reach a satisfactory point for selecting the next block creator (Zhang & Lee, 2020). The Practical Byzantine Fault Tolerance (PBFT) is a less complex and highly practical consensus including five phases of request, pre-prepare, prepare, commit, and reply. PBFT uses three nodes for previously mentioned phases to reach a consistent reply for any request (Zhang & Lee, 2020). Ripple is another consistent consensus that uses a Unique Node List (UNL) of validated nodes for sending transactions through the blockchain nodes. Each validated node sends its transactions to another validated node for verification based on local transaction lists. If the receiving node finds a matching transaction in its local list, the transaction gets validated, and this goes on until the point that at least 50% of validated nodes vote as verified. Also, a screening round will be conducted to reach at least an 80% verification vote for the transaction to be recorded in the ledger (Zhang & Lee, 2020). A comparison of main blockchain consensuses is provided in Table 1 based on fault tolerance, power consumption, scalability, and application (Zhang & Lee, 2020).

Table 1 - Main Consensuses Protocols Comparison (Zhang & Lee, 2020)

Property	PoW	PoS	DPoS	PBFT	Ripple
Type	Probabilistic	Probabilistic	Probabilistic	Absolute	Absolute
Fault Tolerance	50%	50%	50%	33%	20%
Power Consumption	Large	Less	Less	Negligible	Negligible
Scalability	Good	Good	Good	Bad	Good
Application	Public	Public	Public	Permissioned	Permissioned

The aforementioned consensuses along with smart contracts empower blockchains with great capabilities like peer-to-peer data sharing (Huumo et al., 2016). Smart contracts emerged in 1994 as "computerized transaction protocols that execute the terms of a contract to satisfy common contractual conditions, minimize accidental and malicious exceptions, and minimize the need for trusted intermediates" (Yaga et al., 2018, p 32). In a blockchain context, smart contracts are executed by the nodes and are capable of performing non-financial or financial calculations, storing information, exposing properties, sending funds automatically, etc. Smart contracts can provide attestable data, elevate transparency, and build trust between the nodes of a blockchain network by producing the same output based on a determined input (Yaga et al., 2018). After all, after implementing a blockchain, changing or updating the blockchain's networks and data structure is not easy. It is called forking and is divided into soft forking and hard forking (Yaga et al., 2018). In a soft fork, changes are compatible with previous nodes that are not updated; however, in a hard fork, these changes are not compatible with previous nodes, and new blocks that follow the changes will be rejected from the previous ones (Yaga et al., 2018).

2.6.2 Blockchain Challenges

Nonetheless, not every emerging technology is flawless. In some cases, blockchains are limited or miscomprehended. For example, it is believed that blockchains are immutable, though it is not always true since in some situations the most recent published blocks, or the tail blocks, should be replaced while competing for submission (Yaga et al., 2018). This can happen also in the case of 51% attacks in permissionless blockchains. In addition, it is miscomprehended that blockchains are out of control; however, in a permissionless, permissioned, or consortium blockchain, rules, practices, and processes are placed by the publishing nodes, software developers, users, authorities, or stakeholders to govern the way the blockchain works (Yaga et al., 2018). Technology Acceptance Models (TAM) revealed that the quality of the blockchain-based systems is the factor that influences usefulness perception and intention to use the most. They also indicate that benefit perception is in a direct relationship with user acceptance, and User Interface (UI) design affects the ease-of-use perception the most (Shrestha & Vassileva, 2019).

Also, the GDPR-compliancy of blockchain has always been a topic of discussion. The most contradictory notions around the GDPR-compliancy of blockchains started in 2017 (Haque et al., 2021). Most of the blockchain designs use Write-Once/Read-Many approach that lacks the deletion operation which is a fundamental user's right according to the GDPR acts (Kuperberg, 2020). Moreover, GDPR requires the presence of at least one data controller which is a challenge in decentralized blockchains (Haque et al., 2021). In this regard, solutions have been introduced that each may decrease the featured functionalities of blockchains, like off-chain storage. Offline storage is connected to blockchain storage with a specific key management protocol that helps the logical deletion of data by destroying the mentioned key (Haque et al., 2021; Kuperberg, 2020).

Most of them are not GDPR-compliant and store health data off-chain to acquire GDPR-compliancy in some ways. However, the requirements of a blockchain-based GDPR-compliant health data management system are not restricted to this issue. A GDPR-compliant blockchain-based solution must be able to provide public access to immutable, verified, authentic, and non-repudiable evidence showing agreement between the data subject and data controller (Esmel et al., 2021). In this respect, the data controller and data subject need an explicit consent or smart contract for both collecting and processing health data separately which is valid over a certain period of time. The consent should indicate the ID of the data controller and data subject, data collection or processing lifetime, and the ID of data receivers and data processors. Most importantly, the consent should be publicly available for the data subject to control, check, and revise the consent in case of self-interest (Esmel et al., 2021). Also, access should be modifiable, and rights of erasure, withdrawal and violation notification should be taken into account (Esmel et al., 2021).

A model using tree-based ledgers and linear sub-chains seems to overcome these issues by enabling the deletion of data in a linear sub-chain with smart contracts and consensuses without affecting the integrity of the main chain and other blocks (Kuperberg, 2020). Also, controller role issues have been addressed by various methods such as defining participants nodes as controllers, executive actors as controllers, or federated smart contract-based blockchains (Haque et al., 2021).

Another challenge of blockchain technology is its connection with the real world. It is struggling to detect intentionally or unintentionally false-recorded information, especially in cases where blockchain uses pseudonymization techniques (Yaga et al., 2018). However, techniques like Mineable Oracle Contract and Oraclize have tried to address such issues by translating trusted web API data into blockchain readable data (Yaga et al., 2018). Block size growth is also a struggling issue that decreases speed, scalability, and energy efficiency. In this respect, simultaneous on-chain/off-chain storage has been proposed that decreases security instead, like the example in Figure 6 which operates authentications on-chain and stores actual data off-chain (Chen et al., 2019; Truong, et al., 2019). Moreover, blockchains may never be fully shut down due to decentralization, and a dysfunctional blockchain that is prone to attract attackers may not be suitable for historical records. Cybersecurity risks are not removed by using blockchain. A tailor-made cybersecurity plan may be required for each implementation of blockchain due to the context and procedures involved (Yaga et al., 2018).

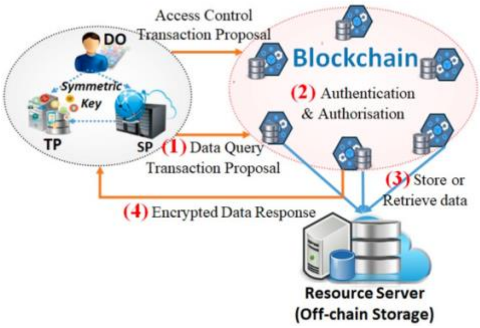


Figure 6 - On-Chain Authentication Off-Chain Storage Personal Data Management (Truong et al., 2019)

Security is one of the most important research topics in the blockchain field. For instance, security challenges like Distributed Denial of Service (DDoS) and Denial of Service (DoS) attacks target yet unpublished blocks which are neither securely published in a block nor become tamper-proof yet (Huumo et al., 2016; Yaga et al., 2018). But attacks are not always from outsiders, they can also occur from insiders. Permissionless blockchains can impose rules but cannot impose ethical codes of conduct. Accordingly, malicious users can operate malicious acts like ignoring transactions, creating secret chains, disrupting information transmission, etc. (Yaga et al., 2018). Another miscomprehension about blockchains is that some individuals think of them as trustless environments that have no trusted third party. Though trust can be built by cryptography, smart contracts, trusted software developments, etc. (Yaga et al., 2018).

Another important challenge of blockchains is resource usage, especially in the models using the PoW consensuses, and inadequate publishing rewards, especially in cryptocurrency cases (Huumo et al., 2016). A fair and logical amount of reward is required for a user to be tempted to contribute to publishing nodes; sometimes the ratio of resource-usage/reward does not make sense or is very volatile (Yaga et al., 2018). Some people think that blockchains support identities through public key infrastructures, which is not true. But they can be used for identity verification and management with the help of outsider processes (Yaga et al., 2018). Although blockchain is innovative and interesting, its implementation should be wisely assessed based on organizational goals (Gatteschi et al., 2018). In the following, applications of blockchain in health data management will be explored along with its cons and pros.

Although blockchain brings innovation, security, and privacy to the health data management system, it adds complexity and has low experimental implementation cases (El-Gazzar & Stendal, 2020). Blockchain is still immature, and its GDPR-compliancy is one of the main concerns that has not been addressed thoroughly. Also, it is important to notice if the blockchain technology is going to be integrated with the current health data management system or is going to redesign it from scratch (El-Gazzar & Stendal, 2020). The main challenges are scalability, regulatory uncertainty, interoperability, irreversible quantum computing, tokenization, integration with the current health system, the accuracy of health data, and cultural adoptions (Yaqoob et al., 2021). Noteworthy to remind that this technology is in its infancy in integration with healthcare and is becoming more and more mature as time passes and challenges may be tackled over time; hence, it is better to understand such systems well in advance.

2.6.3 Blockchain-Based Health Data Management System

Consortium blockchains seem to fit the requirements of personal data management. Hyperledger Fabric is a universal open-source Linux-based blockchain framework that applies consensuses, smart contracts, etc. with the capability of 2000 transactions per second. IT leaders like IBM, Intel, and Huawei use it (Chen, 2019). Blockchain-based medical data protection is a powerful tool for achieving confidentiality, authentication, integrity, and defined access (Zhang et al., 2020).

For instance, the Gem Health Network uses Ethereum's blockchain technology to help all medical stakeholders access the same integrated health information. GHN provides real-time authorized access (McGhin et al., 2019). Estonia, in a practical implementation, uses blockchain technology provided by Guardtime to integrate citizens, health providers, and insurance companies for a better medical treatment performance data retrieval (McGhin et al., 2019). United Arab Emirates (UAE) Ministry of Health and Prevention (MoHAP) has also announced the implementation of blockchain-based health data storing in near future. This system enables reliable and unchangeable track of data and keeps performance records of health workers and organizations. These result in data validation, consistency, process automation, improved sustainability, and operational efficiency (Yaqoob et al., 2021). Swiss also has implemented a blockchain-based system using Hyperledger and a common consensus between the hospitals. This way, the transactions are all stored and traceable without the need for a third party. This Hyperledger is a permissioned blockchain that is based on Switzerland's GIS standard of healthcare messaging format. This has resulted in integrated healthcare actors and provides secure health device tracking (Yaqoob et al., 2021).

There are many different blockchain-based designs for handling health data that their objective use can be compared to each other based on their architecture, data integrity, data sharing, access control, distributed data, patient encryption key, framework, and algorithm (Bodkhe et al., 2020). A 2020 comprehensive study comparing 22 blockchain-based health data management systems based on the aforementioned criteria has mapped the health data transaction in a blockchain framework as illustrated in Figure 7.

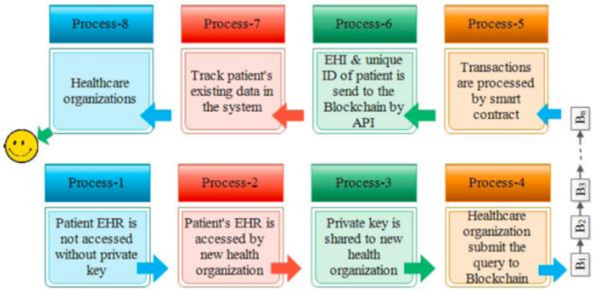

Process-8
Healthcare organizations
Process-7
Track patient's existing data in the system
Process-6
EHI & unique ID of patient is send to the Blockchain by API
Process-5
Transactions are processed by smart contract
B_n
Process-1
Patient EHR is not accessed without private key
Process-2
Patient's EHR is accessed by new health organization
Process-3
Private key is shared to new health organization
Process-4
Healthcare organization submit the query to Blockchain
B_3
B_2
B_1

Figure 7 - Health Data Transactions in the framework of Blockchain (Bodkhe et al., 2020)

In addition, any blockchain-based health data management system should take into account the Health Insurance Portability and Accountability (HIPAA) act of 1996 and GDPR (Alamri et al., 2021). These privacy rules categorize any individually identifiable health information such as demographic and genetic information, that is transmitted or maintained in any form or medium, as Protected Health Information (PHI) (Ahram et al., 2017). HealthChain is one of those designs that is HIPAA compliant. HealthChain considers health data as digital assets secured by encryption and restricted by authorization required access. Transactions are executed based on the smart contracts digitally signed by the patient and third-party data users. HealthChain uses permissioned Hyperledger Fabric and IBM's blockchain for its architecture which is secure, confidential, and scalable (Ahram et al., 2017).

HealthChain is a patient-centered GDPR-compliant blockchain-based framework that acts as a bridge interface between the health stakeholders instead of a health data repository (Hylock & Zheng, 2019). It uses a permissioned blockchain to establish trust, authorization, and interoperability based on HL7 and FHIR. HealthChain breaks down blocks into two types of patients and logs blocks. Logs blocks are not modifiable while patients' blocks are. This way, a full historical immutable record of logs will be available. Patient block handles the patient registration and smart contracts. This block retrieves data from off-storage resources to tackle the data size issue for large medical data sets (Hylock & Zheng, 2019).

Another prominent model for blockchain-based health data management is a three-layered patient-centered architecture that is divided into consortium blockchain, Fast Health Interoperability Resources (FHIR), and Electronic Health Record (EHR) layers (Saha et al., 2021). The bottom layer is the EHR layer, where all independent medical stakeholders lie. The middle layer is a universal standard cloud layer for health information exchange based on Health Level Seven International (HL7) standard. A medical authority is required to manage this layer with the help of APIs. Finally, the upper layer is a hybrid consortium blockchain including regulatory entities (Saha et al., 2021). As illustrated in Figure 8, an end-user application will play the bridge role to connect the users to this system.

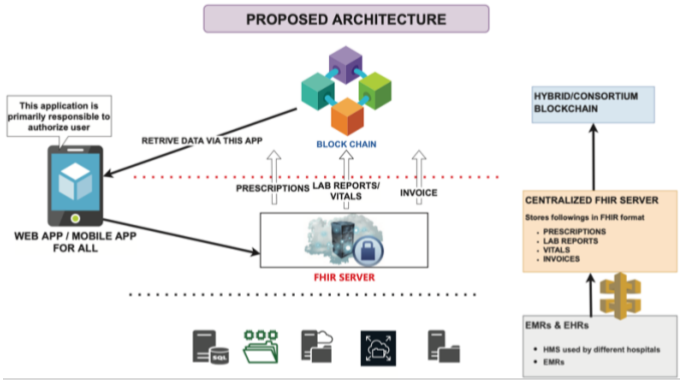

PROPOSED ARCHITECTURE
This application is primarily responsible to authorize user
RETRIVE DATA VIA THIS APP
BLOCK CHAIN
HYBRID/CONSORTIUM BLOCKCHAIN
PRESCRIPTIONS
LAB REPORTS/ VITALS
INVOICE
WEB APP / MOBILE APP FOR ALL
FHIR SERVER
CENTRALIZED FHIR SERVER
Stores followings in FHIR format
PRESCRIPTIONS
LAB REPORTS
VITALS
INVOICES
EMRs & EHRs
HMS used by different hospitals
EMRs
SQL

Figure 8 - Three-Layered Patient-Centered Blockchain-Based Health Data Management (Saha et al., 2021)

Considering the growth of IoT and Big Data use in the medical field, and taking into account the GDPR-compliancy, a 2021 study proposes a six-layered framework for health data management which includes the layers of health IoT data, data privacy, data management, data interoperability, big health data, and stakeholders (Alamri et al., 2021). The Health IoT layer is put as the core layer due to its vast usability of data collection method, and the privacy layer includes the regulatory measures suggested by HIPAA and GDPR. Then, the data management layer must provide secure data sharing between the stakeholders using permissioned blockchains, and the interoperability layer applies the FHIR standard for universal data exchange format (Alamri et al., 2021). Furthermore, the big health data layer applies the sentiment analysis methods for comprehensive use of mass data, and the stakeholder's layer includes all the users such as patients, doctors, researchers, hospitals, etc. (Alamri et al., 2021).

The abovementioned model introduces an Electronic Health Wallet (EHW) working in the data management layer. This EHW allows the patients to access their up-to-date health data in a real-time manner (Alamri et al., 2021). EHW should be able to collect data from IoT devices, EHRs, and patients, and use the measures provided by the data management layer to authorize users, manage accessions, and apply incentives. The main actors in the EHW are patients, health service providers, and third parties (Alamri et al., 2021). Blockchain is the technology that runs the data management layer by providing a historical health data repository and protection measures. This model puts patients in the center of power through smart contracts that control data AddPolicy, UpdatePolicy, HashPolicy, AccessPolicy, and IncentivePolicy. For instance, IncentivePolicy can control incentives or data monetization for data sharing with third parties (Alamri et al., 2021).

One of the most recent published frameworks of health data management that have been proposed based on the real phenomenal outbreak of Covid19 is using blockchain and Artificial Intelligence as complementary technologies to notify the strength and prerequisites of blockchain-based health data management systems (Jabarulla & Lee, 2021). This blockchain-based framework has been designed to tackle the current unagile and provider-centered health data management system. Key stakeholders are defined as patients, providers, payers, pharma, researchers, regulators, and government. Patients are put at the center of the system by being entitled to full access, control, and ownership over their health data. (Jabarulla & Lee, 2021). This patient-centered approach is required to acquire GDPR-compliancy by enabling privacy protection and incentivization. Moreover, it specifically can elevate patient-centered health service provision. However, other stakeholders are also critical and vital (Jabarulla & Lee, 2021).

This is a three-layered framework divided into the blockchain, AI, and decentralized storage layers. These layers are integrated by smart contracts to enable decision-making and accession management for patients (Jabarulla & Lee, 2021). The blockchain layer enables multi-stakeholder interactions and transactions within a patient-centered healthcare system. The decentralized storage layer also uses blockchain technology to enable encrypted peer-to-peer data storage, both on-chain and off-chain. Off-chain storage blockchain saves health data and is not accessible publicly. Saved health data are encrypted and verified by the involved stakeholders and are connected to the on-chain storage blockchain that handles transaction agreements and validation (Jabarulla & Lee, 2021). Finally, the AI layer uses machine learning and deep learning to analyze the health data from the other layers and enhance decision-making (Jabarulla & Lee, 2021).

To go a little deeper into the framework, the blockchain layer consists of provider, user, and training nodes. Provider nodes are any other stakeholders other than the patients, and user nodes are the patients who control their health data with the help of an Ethereum-based smart contract protocol. Eventually, the training nodes are the nodes applying transactions that are connected to the AI layer (Jabarulla & Lee, 2021). Although this blockchain-based approach has advantages like decentralized storing, patient-centricity, GDPR-compliancy, incentivization, anonymization, traceability, and high fault-tolerance, its costs are yet uncertain and it processes minimal transactions which require scalability revisions (Jabarulla & Lee, 2021). In this regard, integrated non-encapsulated cloud blockchains can be used to overcome the scalability problem; however, then, real-time access to data is being impacted (Ismail et al., 2021). Nonetheless, patients are the only true owners of their health data and a balance between the cons and pros of such a system can tackle the barriers of the current health data management system (Jabarulla & Lee, 2021).

SmartMedChain is another blockchain-based example that is designed in three layers. The data layer uses InterPlanetary File System (IPFS) distributed cloud storage to save encrypted health data and EHRs coming from the medical IoT devices and healthcare providers (Majdoubi et al., 2021). This off-chain cloud storage sends stored data's encrypted hash to a blockchain to detect modifications and provides privacy by AES symmetric algorithm. Moreover, each participant node has a local Service registry to save Privacy Offers and establish Privacy Agreements with patients (Majdoubi et al., 2021). The user layer consists of possible stakeholder nodes like patients and healthcare providers. Patient nodes collect data from medical IoT devices and health providers and send it to IPFS storage. Also, patient nodes can run smart contracts and generate blocks for transactions validating and consensus execution. Healthcare provider nodes provide continuous health services and send EHRs to the storage nodes. These nodes can also run smart contracts and execute consensuses. Patient and healthcare provider nodes access blockchain by a web-based application (Majdoubi et al., 2021). The

Control layer handles the registration, certification, blockchains, consensus nodes, access control, and smart contracts.

The registration module is responsible for the identification and verification of new participants, and blockchains are divided into DataChain, ServiceChain, and LogChain which are permissioned blockchains on the Hyperledger Fabric platform (Madjoubi et al., 2021). In this respect, DataChain is responsible for medical IoT data access and publication, ServiceChain is responsible for EHRs access and publication, and Privacy Agreement execution and tracking. Eventually, LogChain is used to keep a tamper-proof record of data accession for further auditing (Madjoubi et al., 2021). The consensus nodes apply PBFT consensus for block propagation and the access control module applies smart contracts in contact with the data layer to handle accessions to data (Madjoubi et al., 2021). Privacy agreement management focuses on the privacy offers provided to a patient by healthcare providers. Privacy offer is a mutual obligation in which healthcare providers clarify the services and data processing procedures and the patients accept to share certain data and participate. Any accepted privacy offer becomes a privacy agreement that will be executed on the blockchain (Madjoubi et al., 2021).

After all, by exploring different methods of blockchain implementation for health data management, one can understand that blockchain has strengths and challenges, and implementation is highly dependent on the nature of regulations and healthcare procedures (Yaqoob et al., 2021).

2.7 Theoretical Framework

This thesis explores blockchain-based health data management systems from a Scandinavian point of view, and Norway, as a Scandinavian country, is the main focus of the study. Since Norway's capital became the 3rd smartest city in 2021 and considering the importance of health data management systems in a Scandinavian smart context, this thesis uses the framework of Smart Asset Development to explore possible blockchain-based developments for the current health data management from a Scandinavian point of view (Choroszewicz & Alastalo, 2021; IMD, 2021; Yigitcanlar et al., 2018).

Health data management systems are considered smart assets of a smart city and technology, policy, and communities are the drivers of its development (Yigitcanlar et al., 2018). In addition, integration, innovation, evidence-based decision making, citizen centricity, sustainability, creativity, effectiveness, efficiency, equality, entrepreneurialism, citizen engagement, openness, resiliency, and technology savviness are the dimensions of its smartness; however, not every smart asset entails all the aforementioned dimensions (Gil-Garcia, as cited in Ismagilova et al., 2019). Accordingly, Figure 9 depicts the framework for a blockchain-based development of the current health data management systems in a Scandinavian context.

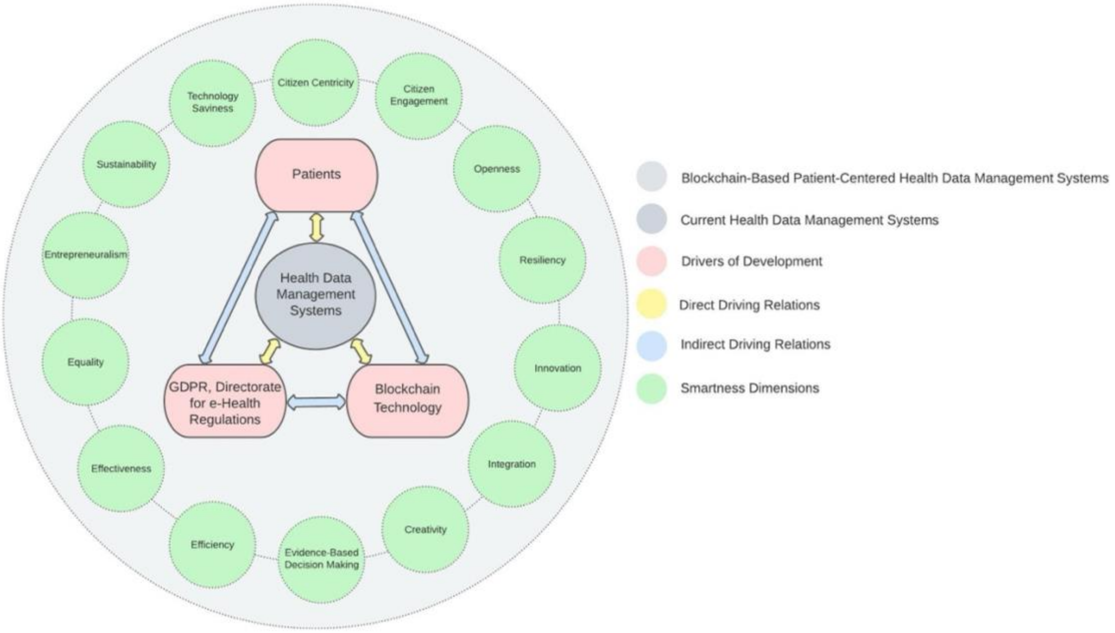


Figure 9 - Health Data Management Development Framework (Adapted From Yigitcanlar et al., 2018; Gil-Garcia et al., 2016)

Figure 9 is a Scandinavian customized version of the smart asset development framework introduced by Yigitcanlar et al. (2018). Since GDPR and regulations of the Directorate for e-health are the imposing rules for handling health data management in Norway, they are put instead of policy (EU, 2022; Vigot & Bussche, 2017; Yigitcanlar et al., 2018). Moreover, considering the importance of a patient-centered approach in handling health data management in Scandinavia, patients are put instead of communities (IMD, 2021; Laurini, 2020; Yigitcanlar et al., 2018). Finally, due to the high-tech infrastructural richness in Norway and considering the importance of creative solutions for smartness sustainability, blockchain technology, as a disruptive technology, is put instead of technology (IMD, 2021, Laurini, 2020, Yigitcanlar et al., 2018).Based on this framework, a thorough literature review, data collection, and data analysis has been conducted which are explained in the next section.

First, the systematic literature review explores the concepts of health data management, blockchain, and regulations. Second, the data collection extracts meaningful patterns of data from scholarly publications and tweets that discuss these concepts. Third, the data analysis mines the collected data for semantic analysis. Methodologies used in each of these steps are explained in the following.

3 Methodology

This thesis used the systematic literature review guidelines introduced by Webster & Watson (2002) for the review process, and the inductive approach and qualitative methods introduced by Bryman & Bell (2011) for study design, approach acquisition, and reliability, validity, and generalizability evaluation.Moreover, the RStudio programming tool and Twitter platform were chosen as the data collection and analysis assistive tools regarding their capabilities in providing a huge amount of structured and unstructured data and processing large sets of data (Rani & Rani, 2016). Descriptive analysis methods provided by these tools help the understanding of interrelations between the data sets and the transformation of these data sets to actionable knowledge for decision-makers (Rani & Rani, 2016).

In the following, the review process, approach, data collection, data analysis, reliability and validity, and limitations will be explained.

3.1 Review Process

This thesis is an exploratory study of health data management systems in Scandinavia. Hence, the scope of the study is Scandinavia, and the main focus is on Norway as a sample of Scandinavian countries. A systematic approach is used for the search process of prominent scholarly articles based on the guidelines introduced for writing systematic literature reviews (Webster & Watson, 2002). The review process started with defining high-ranked databases suitable for the topic of interest. Accordingly, Google Scholar, IEEE, Web of Science, and PubMed were selected to sufficiently cover both technology and health fields. To cover the driving factors of policy and technology in the grounding smart development framework, and due to the research question, keywords were identified as “blockchain”, “data management system”, “data sharing system”, “health”, “GDPR”, “Norway”, “Nordic”, and “Scandinavia”. The whole process consists of 52 times of search queries including different combinations of keywords with the operators of “AND” and “OR” in the aforementioned databases, as shown in Table 2.

For this purpose, Google Scholar, IEEE, Web of Science, and PubMed were searched, respectively, 18, 12, 11, and 11 times with the highest number of records of 343,000 and the lowest of 0. The total number of search records is 627,633 hits.

Table 2 - Search Queries Summary

Engine	Search Query	No of Records	Year
Google Scholar	["blockchain"] AND ["data management system" OR "data sharing system"]	3,590	No Filter
	["blockchain"] AND ["data management system" OR "data sharing system"] AND ["Norway" OR "Nordic" OR "Scandinavia"]	190	2015-2022
IEEE	"blockchain"	10,865	No filter
	("blockchain") AND ("data management system" OR "data sharing system") AND ("health")	6	2020-2022
Web of Science	("extended reality" OR "virtual reality" OR "augmented reality") AND ("depth perception") AND ("stereopsis" OR "stereoscopic" OR "vision")	136	2015-2022
	"data management system" OR "data sharing system"	1,394	No Filter
PubMed	"blockchain"	744	2020-2022
	("blockchain") AND ("data management system" OR "data sharing system") AND ("health")	148	2020-2022

To extract grounding and foundational articles around the topic the queries started from general areas of blockchain, data management, and data sharing with no filter for the publication year. However, citation of the articles and authority of the journals were considered the main factor in selection. Queries were then filtered three times each to extract the records that belong to the periods of 2000-2022, 2015-2022, and 2020-2022 separately. These periods were selected to identify respectively the grounding articles, recent top articles, and state-of-the-art articles around the topic. Finally, after three rounds of filtration, 187 articles and proceedings were eligible based on the considered inclusion and exclusion criteria. The first round of filtration started with reading the title of the article to see if it is related at all, and in the second round, the publication year, citation, and authority of the article were taken into account to increase the validity and authority. Ultimately, the abstract and conclusion of the articles from the second round were read to indicate the relevance and eligibility of the publication. A flowchart of the whole search process is illustrated in Figure 10.

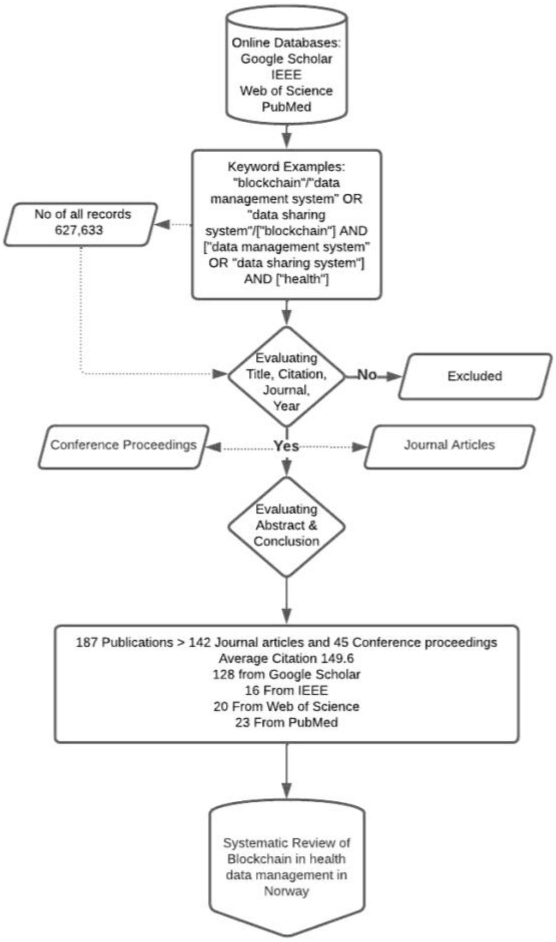

Online Databases:
Google Scholar
IEEE
Web of Science
PubMed
Keyword Examples:
"blockchain"/"data management system" OR "data sharing system"/["blockchain"] AND ["data management system" OR "data sharing system"] AND ["health"]
No of all records
627,633
Evaluating Title, Citation, Journal, Year
No
Excluded
Conference Proceedings
Yes
Journal Articles
Evaluating Abstract & Conclusion
187 Publications > 142 Journal articles and 45 Conference proceedings
Average Citation 149.6
128 from Google Scholar
16 From IEEE
20 From Web of Science
23 From PubMed
Systematic Review of Blockchain in health data management in Norway

Figure 10 - Search Process Flowchart

A summary of the chosen articles is illustrated in Table 3.

Table 3 - Selected Publications Summary

Year	No of Records	Portion	No of Selected Publications	Citation Average
1967-2022	627,633	100%	187	149.6
2015-2022	92,755	95%	179	150.4
2020-2022	39,947	55%	104	43.1

As it is shown in Table 3, the average citation of all the articles is 149.6 which is at an acceptable level, and the average citation for the articles from 2020 up to now is 43.1 which shows the importance of the selected articles. It is crystal clear that research around the combination of blockchain and data management is increasing greatly due to the fact that 95% of the publications belong to the last 8 years and 55% of them belong to 2020 up to now. Although, only a very few numbers of them are concentrating on health data management systems, and even less are concentrated on Nordic countries.

3.2 Approach

Since this thesis aims to explore health data management systems from a Scandinavian point of view to draw out the challenges and proposition of blockchain-based data subject-centered solutions, the inductive approach has been chosen to present a possible conceptualized model after exploration, data collection, and analysis (Bryman & Bell, 2011). The inductive approach helps generalization and reasoning after gathering a rich set of grounding data (Bryman & Bell, 2011). In this regard, 187 high-quality publications regarding data management, data sharing, e-health, blockchain, security, privacy, GDPR, Scandinavia, Nordic, and Norway were extracted through a systematic process introduced by Webster and Watson (2002), as explained in Section 2.1. This way, a thorough literature exploration regarding the topic was made to extract the notions of scholars around the topic.

In addition, a qualitative interpretive approach was taken into account to mine 193 pdf files, including 187 publications and 6 regulatory pdf files, along with Twitter data with the Rstudio programming tool to draw out common scholarly and user-centered beliefs around the topic (Bryman & Bell, 2011). The qualitative approach helps the interpretation of the text, speech, voice, gesture, etc. (Bryman & Bell, 2011). In this regard, pdf files were categorized into four main sets of documents each concentrating on either blockchain, data management, e-health and blockchain, or security, privacy, and GDPR. This main categorization was done by investigation of the keywords, abstracts, and conclusion of the documents. In addition, tweets were collected via Twitter over 30 days including tweets separately from Norway, Finland, Sweden, and Denmark consisting of either data management, e-health and blockchain, or security, privacy, and GDPR words and hashtags. This way, a collection of tweets was made available and categorized into seven main categories of Norway, Denmark, Sweden, Finland, data management, e-health and blockchain, and security, privacy, and GDPR.

3.3 Data Collection

For data collection and data mining, Twitter and Rstudio were used due to their capabilities in the extraction of opinions and sentiment analysis. This helps to achieve a general understanding of the scholars' emotions and beliefs around blockchain technology, GDPD, and regulations of the

Directorate for e-health (Rani & Rani, 2016). In addition, to cover the point of view of the patients, as the other driving factor in the smart asset development framework, tweets were collected to achieve a general understanding of their emotions and beliefs around the driving factors. In this regard, two methods were mixed; first, top-quality publications were extracted, as explained in Section 2.1., and then regulatory documents regarding GDPR and e-health regulations were added to the selection. The first step provided 193 documents. Second, the Rstudio programming tool was used to grab data from both the 193 pdf documents and the Twitter website. To ensure privacy, collected data from Twitter were anonymized by removing usernames and no names are mentioned in the results. Also, the main sources of data are saved on an internal hard drive with encrypted access for a maximum 5-year period to provide an archived reference for the thesis.

To collect and clean the data, a Twitter Research and a Google Cloud Platform account were created to acquire Twitter API and Google API tokens to be able to use packages and libraries of rtweet, geolocation, writexl, wordcloud, wordcloud2, tm, lsa, nlp, rjava, corpus, tables, dplyr, readr, devtools, ROAuth, ggplot2, stringr, etc. along with some functions like tm_map, SentimentAnalysis, removewords, stopwords, stemdocumen, stemcompletion, termdocumentmatrix, documenttermmatrix, removepunctuation, removenumbers, stripwhitespace, analyzeSentiment, etc. (Rani & Rani, 2016; Rstudio Community, 2022).

For this purpose, 193 documents were categorized into four groups based on their keywords, abstract, and conclusion. These four groups were blockchain, e-health and blockchain, data management, and security, privacy, or GDPR. This way it is possible to conduct a thorough document analysis via Natural Language Processing tools that explore emotions, beliefs, negativity, positivity, and neutrality in texts like SentimentAnalysis, and even correlational relationships between a set of documents like Latent Semantic Analysis (Rani & Rani, 2016; Rstudio Community, 2022).

The same process of collecting and cleaning the data was conducted for tweets while tweets were categorized into twelve subcategories with three main categories of data management, e-health and blockchain, and security, privacy, or GDPR each consisting of a subgroup of Norway, Finland, Sweden, Denmark. This way, it is possible to analyze the collected tweets based on text, keywords, hashtags, and locations. Ultimately, different visualization and exporting methods were used to export the analyzed data from Rstudio (Rani & Rani, 2016).

3.4 Data Analysis

The most frequent words and concepts of each category were extracted with the help of the RStudio tools to investigate the most important concerns of each category. Moreover, the total emotion around each category and the relationship between the categories were extracted to investigate the common beliefs of scholars around the driving factors. For instance, to extract the most important concepts in the category of e-Health and blockchain documents, as the technology factor, and its relationship with the category of the documents that include GDPR, security, and privacy, as the policy factor. For this purpose, wordcloud and wordcloud2 were used to extract the most frequent concepts and words both in the pdf texts and in the tweets (Moumen & Mejjad, 2021; Rstudio Community, 2022).

For example, the below lines of Rstudio codes are used to collect, clean, prepare, and mine the e-Health and Blockchain data from the publications to conduct a single cloud of words analysis.

```
setwd("~/Documents/USN/MIS/4th Semester - Master Thesis/Data Collection/Twitter Data/PDF References
Categorized/eHealth - Blockchain")
PDF_HB_List<- list.files(pattern = "pdf$")
PDF_Text_HB<-lapply(PDF_HB_List, pdf_text)
PDF_Text_HB_O <- read_csv("Documents/USN/MIS/4th Semester - Master Thesis/Data Collection/Twitter
Data/Reference Text CSV/PDF_Text_HB_O.csv")
capture.output(PDF_Text_HB, file = "PDF_Text_HB_O.csv")
PDF_Text_HB_DF <- as.data.frame(t(stri_list2matrix(PDF_Text_HB)))
colnames(PDF_Text_HB_DF) <- unique(unlist(sapply(PDF_Text_HB,names)))
str(PDF_Text_HB_DF)
PDF_Text_HB_DF_Cleaned <- tolower(PDF_Text_HB_DF)
PDF_Text_HB_DF_Cleaned <- removePunctuation(PDF_Text_HB_DF_Cleaned)
PDF_Text_HB_DF_Cleaned <-removeNumbers(PDF_Text_HB_DF_Cleaned)
PDF_Text_HB_DF_Cleaned <- stripWhitespace(PDF_Text_HB_DF_Cleaned)
PDF_Text_HB_DF_Cleaned <- removeWords(PDF_Text_HB_DF_Cleaned,stopwords("english"))
PDF_Text_HB_DF_Cleaned <- stemDocument(PDF_Text_HB_DF_Cleaned)
PDF_Text_HB_DF_Cleaned_TDM <- Corpus(VectorSource(PDF_Text_HB_DF_Cleaned))
PDF_Text_HB_DF_Cleaned_TDM <- TermDocumentMatrix(PDF_Text_HB_DF_Cleaned_TDM)
PDF_Text_HB_DF_Cleaned_TDM_Matrix <- as.matrix(PDF_Text_HB_DF_Cleaned_TDM)
WPDFHB <- sort(rowSums(PDF_Text_HB_DF_Cleaned_TDM_Matrix),decreasing=TRUE)
WPDFHBF <- data.frame(word = names(WPDFHB),freq=WPDFHB)
wordcloud2(data=WPDFHBF, size=1.6, color='random-dark')
```

Plus, to acquire a holistic approach to the attitudes of the scholars and patients, the Natural Language Processing (NLP) and Sentiment Analysis tools from Rstudio were used. This way, the neutrality, negativity, or positivity of the pdf texts and tweets about the driving factors are possible to extract (Kanakaraj et al., 2015; Rstudio Community, 2022). The SentimentAnalysis tool uses Harvard-IV psychological dictionary, Henry's finance-specific dictionary, Loughran McDonald's finance-specific dictionary, and polarity words dictionary to mine opinions and attitudes of the texts (Rstudio Community, 2022). In addition, the Latent Semantic Analysis (LSA) is used to create a well-connected representation of the publications and tweets. LSA creates a latent semantic space of concepts and words from publications and tweets, and then links these spaces and extracts the coherence of the concepts in them (Foltz et al., 1998; Rstudio Community, 2022).

3.5 Reliability, Validity, Generalizability

In terms of validity, reliability, and generalizability, this thesis uses the trustworthiness and authenticity criteria introduced in Bryman & Bell (2011). In this regard, trustworthiness is divided into four criteria of credibility, transferability, dependability, and confirmability, and authenticity is divided into five criteria of fairness, ontological authenticity, educative authenticity, catalytic authenticity, and tactical authenticity (Bryman & Bell, 2011).

Regarding credibility, different scholarly articles from 1967 to 2022 that are published in high-quality journals or conferences were explored by keywords, abstracts, and conclusions to explore the development of data management systems and extract different notions regarding blockchain-based data subject-centered solutions. Regarding transferability, Scandinavia, Nordic, Norway, Finland, Sweden, and Denmark were part of the keywords of either literature review, document mining, or Twitter mining to be able to create a rich Scandinavian context with a slightly more focus on Norway as a sample of Scandinavia. To ensure dependability, all the 187 publications are listed in detail in a research matrix Excel file, all 193 mined documents are stored in a Rstudio saving file, and all the 30 days tweets are stored separately by country, category, and date in a Rstudio saving file. Regarding confirmability, although this thesis aims to conceptualize a blockchain-based data subject-centered health data management system, it is tried to extract both positive and negative notions around the topic to ignore any biased evaluation.

To ensure fairness in authenticity, data were collected and mined from the point of view of scholars, experimental real cases, governmental regulators, and users on Twitter. Moreover, Twitter as an elite social media helps ontological and educative authenticity as it embraces participants of any kind in either identified or anonymized manner, this way users share their notions with others freely and voluntarily.

3.6 Limitations

One of the limitations of this thesis is the small number of scholarly articles relating to Scandinavia, Norway, Sweden, Denmark, and Finland. Moreover, another limitation of this thesis is the lack of data on all of the direct and indirect driving relationships, leading to partial analysis that may impact the generalization of the analysis output. In addition, catalytic and tactical authenticity are the limitations of this study as it is almost impossible that the findings of this thesis impact the anonymous users. Time shortage is another limitation of this study that restricts the data collection the most. Otherwise, if there was no lack of time interviews could also be conducted to gather the point of view of the other stakeholders as well. Finally, the ethical limitation of this thesis is the result communication. Since this thesis uses publicly available data from publications and Twitter, no individual participation request or informed consent was required, but transferring the results to the anonymized users is impossible. A thesis description was provided to Twitter and Google to access data, authorize the identity of the researcher, and ensure ethical considerations.

4 Results

The results section is divided into three subsections of collected data, attitudes, and relations. The collected data section relates to the amount of collected data and its relationship with the health data management development framework. The attitudes section relates to the most frequently discussed topics and concepts in the publications and tweets and their neutrality, positivity, and negativity about the health data management development and its driving factors. Finally, the relationship section relates to the direct and indirect relationship of the driving factors with the health data management development framework.

4.1 Collected Data

The first phase of data collection started with collecting scholarly publications based on the review process explained in section 3.1., and the results are shown in Table 4. Publications are divided into data management and its development driving factors based on the smart asset development framework explained in section 2.7. The second phase of data collection was about collecting tweets from users located in Norway, Sweden, Denmark, and Finland based on the process explained in section 3.3., and the results are shown in Table 5. Tweets also are divided into data management and its development driving factors from the smart asset development framework. This division has been made to have enough data both about the data management as the smart asset itself and about its development driving factors which in the scope of this thesis are blockchain, patients, and GDPR and regulations of the Directorate for e-Health. Locations are also selected to represent Scandinavia.

Table 4 - Number and Dispersion of Collected Articles Based on Keywords, Abstract, and Conclusion

Number of Publications	Keywords, Abstracts, and Conclusion of the publications				
	Smart Asset	Smart Asset Development Driving Factors			
		Technology		Policy	
	Data Management	Blockchain	e-Health/Blockchain	Security/Privacy/GDPR/Directorate for e-Health	Total
	19	163	71	52	193

The focus of this study is on a blockchain-based patient-centered development for the current health data management systems in Scandinavia. Accordingly, table 4 indicates that keywords, abstract, and conclusion of 19 publications were especially regarding data management, which is a smart asset in our development framework, 163 were including blockchain and 71 were including e-health and blockchain which are the technology factor that drives our smart asset development framework. Also, 52 publications were especially regarding security, privacy, GDPR, and regulations of the Directorate for e-Health in Norway. Finally, 193 publications were collected and divided into the smart asset and its development driving factors. However, there is still another development driving factor which is the users that influence the smart asset development.

In this regard, Table 5 shows the tweets of the tweeter users that are considered as users or patients of the smart asset. Table 5 shows the results of 30 days of Twitter data collection that has resulted in the collection of 17,477 tweets. A general view of the results indicates that users are the most concerned about security, privacy, GDPR, and related topics. 70.8% of the collected tweets are in this category.

Table 5 - Number and Dispersion of Tweets Based on Location and Keywords

Keywords of the Tweets				
Location of Tweets	Smart Asset	Smart Asset Development Driving Factors		
		Technology	Policy	
	Data management	e-Health/Blockchain	Security/Privacy/GDPR	Total
Norway	100	2,752	6,321	9,173
Sweden	52	1,127	3,781	4,960
Denmark	2	55	284	341
Finland	26	993	1,984	3,003
Total	180	4,927	12,370	17,477

Technology is the second most concerned topic with a 28.2% proportion and data management is the least concerned one with a 1% proportion of the total tweets. Interestingly, this pattern is the same even in a country-based division. In addition, Norwegian users have been the most active users around these topics with a 52.5% proportion, and Sweden, Finland, and Denmark are the next ones, respectively, with a 28.4%, 17.2%, and 1.9% proportion of tweets.

4.2 Most Discussed Concepts and Attitudes

To acquire a holistic approach toward the most discussed topics and attitudes of the publications and tweets about the data management and its development driving factors, the texts of collected pdf publications and tweets were extracted, cleaned, and mined with the help of word and opinion mining tools in Rstudio, as explained in section 3.3. and 3.4. Each of the Cloudwords below shows the most frequent topics, words, and concepts discussed in the pdf and tweets texts, and the tables show the opinions and attitudes. These most frequent concepts and attitudes are extracted from the publications about data management and its technological and regulative driving factors, which are blockchain and security and privacy-related documents. Blockchain documents are mined both solely and in integration with e-health documents, and GDPR and regulations of the Directorate for e-Health documents are mined as security and privacy policies. Tweets are mined in three categories of data management, e-health and blockchain, and security, privacy, and GDPR.

4.2.1 Most Discussed Concepts

Figure 11, and Table 6 are showing the most discussed concepts in publications concerning data management. Some non-related words that had not been removed in the first cleaning phase are visualized in the cloud word that required a second cleaning phase. Data management is the smart asset that this thesis is trying to explore its development based on a blockchain-based patient-centered approach.

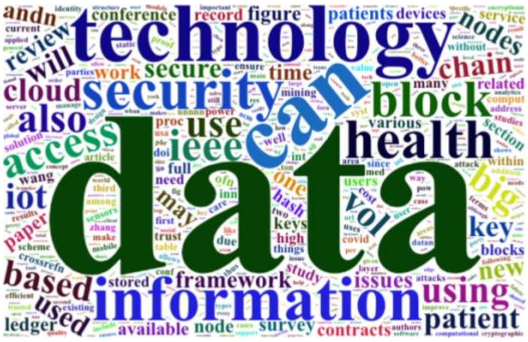


Figure 11 - The Most Discussed Topics in Data Management Publications

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 6 - The Top 10 Discussed Topics in Data Management Publications

Concept	Blockchain	Healthcare	Technology	Information	Security	Network	Smart	Applications	Access	Privacy
Occurrence	3,014	898	750	579	536	526	444	413	411	405

Figure 12, and Table 7 are showing the most discussed concepts in publications concerning blockchain. The same process for the second and final round of data cleaning was done for this cloud word and table too. Blockchain is one of the health data management development driving factors in the context of this thesis.

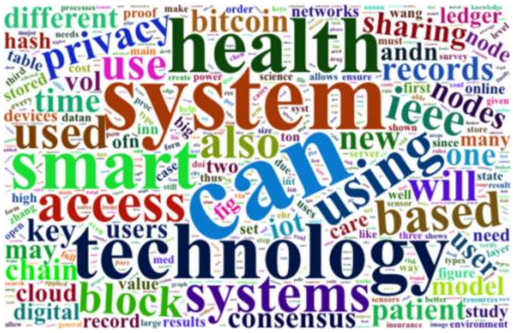


Figure 12 - The Most Discussed Concepts in Blockchain Publications

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 7 - The Top 10 Discussed Concepts in Blockchain Publications

Concept	Data	Technology	Smart	Information	Network	Security	Access	Research	Managemen t	Transaction
Occurrence	17,983	5,200	4,743	4,454	4,271	4,127	3,638	3,374	3,012	2,725

Figure 13, and Table 8 are showing the most discussed concepts in publications concerning blockchain and e-health. The same process for the second and final round of data cleaning was done for this cloud word and table too. Blockchain as the technological development driving factor was explored also in integration with e-health to explore the most discussed concepts when these two topics are tied together.

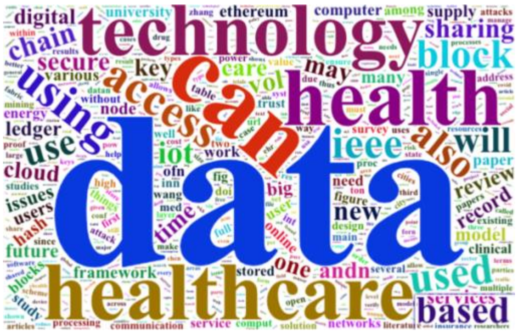


Figure 13 - The Most Discussed Concepts in Blockchain and e-Health Publications

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 8 - The Top 10 Discussed Concepts in Blockchain and e-Health Publications

Concept	Technology	Smart	Security	Network	Information	Access	Application	Management	Privacy	Transaction
Occurrence	2,604	2,571	2,125	2,019	1,986	1,676	1,548	1,442	1,363	1,306

Figure 14, and Table 9 are showing the most discussed concepts in publications concerning security, privacy, GDPR, and regulations of the Directorate for e-Health. The same process for the second and final round of data cleaning was done for this cloud word and table too. Security and privacy-related regulations like GDPR and the Directorate for e-Health are the policy factor in health data management development.

Figure 14 - The Most Discussed Concepts in Security and Privacy Publications

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 9 - The Top 10 Discussed Concepts in Security, Privacy, GDPR, and Directorate for e-Health Publications

Concept	Blockchain	Data	Healthcare	Smart	Technology	Information	Security	Network	Access	Applications
Occurrence	8,612	6,040	2,159	2,143	1,982	1,612	1,599	1,566	1,316	1,173

After extracting the most discussed concepts in the publications that are related to the smart development of health data management systems based on a blockchain-based approach, this thesis is extracting the most discussed concepts between the users to acquire a holistic approach toward the user-centricity or patient-centricity in blockchain-based developments of health data management systems. In this regard, Figure 15 and Table 10 are showing the most discussed concepts when users are talking about data management on Twitter. Here also, a second and final round of cleaning was required to remove non-related or leftover redundant words.

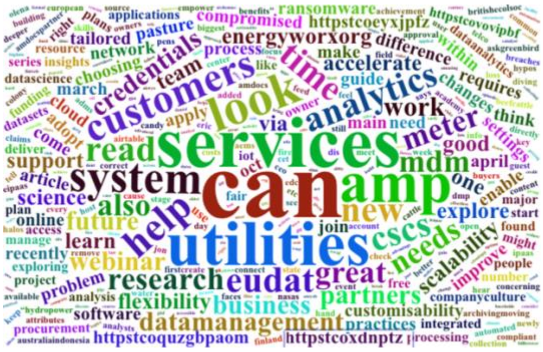


Figure 15 - The Most Discussed Concepts in Tweets About Data Management

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 10 - The Top 10 Discussed Concepts in Tweets About Data Management

Concept	Services	Utilities	Systems	Customer	Help	Analytics	New	Time	Needs	Research
Occurrence	16	14	10	9	9	8	8	8	7	7

Furthermore, Figure 16 and Table 11 are showing the most discussed concepts when users are talking about the e-health and blockchain on Twitter. Here also, a second and final round of cleaning was required to remove non-related or leftover redundant words.

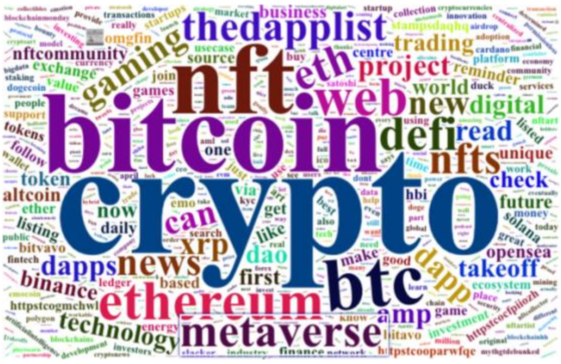


Figure 16 - The Most Discussed Concepts in Tweets About e-Health and Blockchain

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 11 - The Top 10 Discussed Concepts in Tweets About e-Health and Blockchain

Concept	Crypto	Bitcoin	NFT	Ethereum	Web	Metaverse	Decentralized	News	Gaming	Project
Occurrence	644	465	409	223	220	186	163	161	148	121

Finally, Figure 17 and Table 12 are showing the most discussed concepts when users are talking about security, privacy, and GDPR on Twitter. Here also, a second and final round of cleaning was required to remove non-related or leftover redundant words. This phase of cleaning was a bit different. Since users were impacted by the ongoing war between Russia and Ukraine, the tweets about security included war-related concepts that were removed to purify the data. As illustrated in Figure 17, you can see the impact, but Table 12 shows the top 10 purified discussed concepts.

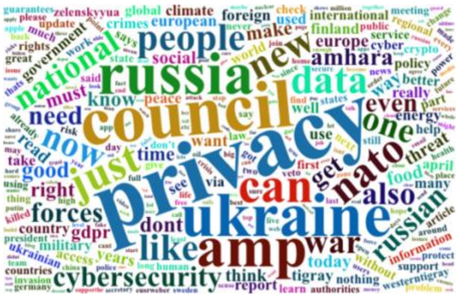


Figure 17 - The Most Discussed Concepts in Tweets About Security, Privacy, and GDPR

After a final round of cleaning data, the top 10 discussed concepts are indicated in the table.

Table 12 - The Top 10 Discussed Concepts in Tweets About Security, Privacy, and GDPR

Concept	Data	New	People	Cybersecurity	Need	Right	GDPR	Threat	Social	Think
Occurrence	313	308	297	223	208	160	146	142	140	139

As a summary of text mining in the publications and tweets, Table 13 indicates the most discussed concepts based on the health data management development framework.

Table 13 - Summary of the Top 10 Discussed Concepts From the Publications and Twitter Sources (Based on the Framework)

Top 10 Discussed Concepts					
Source		Smart Asset	Development Driving Factors		
			Technology	Policy	Communities
		Data Management	Blockchain/e-Health	GDPR/e-Health Directorate/Security/Privacy	Patients
	Publications	Blockchain	Technology	Blockchain	No Direct Data
		Healthcare	Smart	Data	No Direct Data
		Technology	Security	Healthcare	No Direct Data
		Information	Network	Smart	No Direct Data
		Security	Information	Technology	No Direct Data
		Network	Access	Information	No Direct Data
		Smart	Application	Security	No Direct Data
		Applications	Management	Network	No Direct Data
		Access	Privacy	Access	No Direct Data
		Privacy	Transaction	Applications	No Direct Data
	Twitter	Services	Crypto	Data	No Direct Data
		Utilities	Bitcoin	New	No Direct Data
		Systems	NFT	People	No Direct Data
		Customer	Ethereum	Cybersecurity	No Direct Data
		Help	Web	Need	No Direct Data
		Analytics	Metaverse	Right	No Direct Data
		New	Decentralized	GDPR	No Direct Data
		Time	News	Threat	No Direct Data
		Needs	Gaming	Social	No Direct Data
		Research	Project	Think	No Direct Data

A view over Table 13 reveals that after 3 times of publication text mining, the concepts of technology, information, security, network, smart, applications, and access have appeared 3 times in the most discussed topics. The blockchain, healthcare, and privacy concepts each have appeared 2 times, and management, data, and transaction concepts each 1 time. Interestingly, after 3 times of tweet mining, none of the most discussed concepts on Twitter have appeared with the same name; however, they are highly related to some of the other concepts. Their relation will be discussed in the discussion section.

4.2.2 Attitudes

To identify the sentiment of the publications and tweets, a sentiment, negativity, positivity, and polarity analysis has been made based on the smart asset development framework. Table 14 is a summary of the mined publications and tweets.

Table 14 - Summary of Sentiment Analysis From the Publications and Twitter Sources (Based on the Framework)

Sentiment Analysis of Attitudes						
Source	Smart Asset			Development Driving Factors		
				Technology	Policy	Communities
	Data Management			Blockchain/e-Health	GDPR/e-Health Directorate/Security/Privacy	Patients
	Publications	Word Count	121,652	494,131	71,919	No Direct Data
		NegativityIG	0,05847862	0,057411292	0,076636743	No Direct Data
		PositivityIG	0,163541296	0,172581769	0,213943785	No Direct Data
		SentimentIG	0,105062674	0,115170477	0,137307042	No Direct Data
		NegativityQDAP	0,035730617	0,034460107	0,053484551	No Direct Data
		PositivityQDAP	0,112967389	0,118324755	0,155297153	No Direct Data
		SentimentQDAP	0,077236772	0,083864648	0,101812602	No Direct Data
	Twitter	Word Count	3,392	63,527	175,342	No Direct Data
		NegativityIG	0,04852878	0,052658364	0,085900945	No Direct Data
		PositivityIG	0,223531406	0,152597192	0,228006948	No Direct Data
		SentimentIG	0,175002626	0,099938827	0,142106003	No Direct Data
		NegativityQDAP	0,021457515	0,032472618	0,061337936	No Direct Data
		PositivityQDAP	0,163212573	0,114466818	0,17004787	No Direct Data
		SentimentQDAP	0,141755057	0,0819942	0,108709934	No Direct Data

This mining counts words and extracts the sentiment of the text. This sentiment analysis then examines the neutrality, negativity, and positivity of the text. In this part of mining also, another phase of cleaning was required due to some invalid results that were created out of erroneous records. Those records were removed. SnetimentGI represents the attitudes based on the Harvard-IV psychological dictionary, and SentimentQDAP is based on the polarity dictionary.

A general view over Table 14 indicates that 687,702 words from the publications and 242,261 words from tweets have been mined which is an aggregate of 929,963 mined words. As colored in the table, all the sentiments are positive; however, there are variations. The table will be explained fully in the discussion section.

4.3 Relationships

A Latent Semantic Analysis (LSA) was conducted to extract the correlations between the elements of the smart development framework. In the context of this thesis, data management is the smart asset, blockchain represents the technological driving factor, and security, privacy, GDPR, and regulations of the Directorate for e-Health represent the policy factor. Furthermore, Patients represent the community factor which is the Twitter users in our data collection.

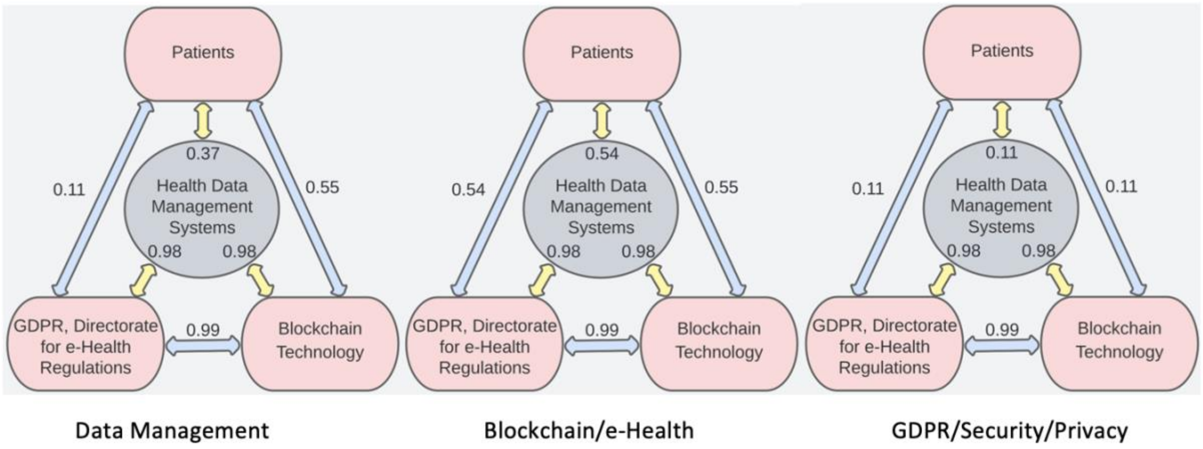
A latent semantic space was created to be able to examine the correlations between the elements based on the Pearson method. The publications and tweets were categorized into seven subsets of documents to explore the correlation between them. Table 15 indicates the correlation between the data management, and its blockchain-based patient-centered development driving factors from a Scandinavian point of view.

Each correlation between each two of these seven subset documents represents one of the direct or indirect relationships between the elements of the smart asset development. As illustrated previously in Figure 9, there are three two-way relationships between the smart asset development driving factors and three two-way relationships between the smart asset and its three development driving factors. Correlations of more than 0.5 are colored in the table.

Table 15 - Correlations Between the Elements of Blockchain-Based Patient-Centered Health Data Management (From a Scandinavian Point of View)

			Source						
			Publications				Twitter		
		(Correlation Matrix)	Data Management	Blockchain/e-Health	GDPR/e-Health Directorate/Security/Privacy	Blockchain	Data Management	Blockchain/e-Health	GDPR/e-Health Directorate/Security/Privacy
Source	Publications	Data Management	1,00	0.98	0.98	0.97	0.37	0.54	0.11
		Blockchain/e-Health	0.98	1,00	0.99	0.99	0.36	0.55	0.11
		GDPR/e-Health Directorate/Security/Privacy	0.98	0.99	1,00	0.99	0.37	0.54	0.11
		Blockchain	0.97	0.99	0.99	1,00	0.41	0.51	0.12
	Twitter	Data Management	0.37	0.36	0.37	0.41	1,00	0.12	0.19
		Blockchain/e-Health	0.54	0.55	0.54	0.51	0.12	1,00	0.15
		GDPR/e-Health Directorate/Security/Privacy	0.11	0.11	0.11	0.12	0.19	0.15	1,00

An indication of correlation of elements is illustrated below. Figure 18 shows three different correlation groups that each represent the sets of correlations when users talk about either data management, blockchain and e-Health, or security, privacy, and GDPR.


Patients
0.37
Health Data Management Systems
0.11
0.55
0.98
0.98
GDPR, Directorate for e-Health Regulations
0.99
Blockchain Technology
Patients
0.54
Health Data Management Systems
0.54
0.55
0.98
0.98
GDPR, Directorate for e-Health Regulations
0.99
Blockchain Technology
Patients
0.11
Health Data Management Systems
0.11
0.11
0.98
0.98
GDPR, Directorate for e-Health Regulations
0.99
Blockchain Technology

Data Management Blockchain/e-Health GDPR/Security/Privacy

Figure 18 - Correlation Groups Divided by Users' Discussed Concepts

A general view of the correlations reveals that when the users discuss blockchain and e-Health, their notions correlate the most with the health data management smart development framework extracted from the publications. However, when they discuss data management correlations decrease a bit, especially with regard to policy factors. Finally, when they discuss GDPR, security, and privacy their notions are somehow not correlated at all to the framework. It is noteworthy to remind, correlations that come from health data management systems and their technology, and policy factors are calculated from the discussed topics in the publications. This differentiation of correlations will be explained in detail in the discussion section.

5 Discussion

The contribution of this thesis is to explore the common belief around blockchain-based developments for the current health data management systems. This exploration is patient-centered and includes real anonymized data from Norwegian, Swedish, Danish, and Finnish users. Moreover, the regulations of the Directorate for e-Health of Norway have been surveyed to consider possible conceptual blockchain-based patient-centered solutions in a Scandinavian smart asset development context.

5.1 Blockchain-Based Health Data Management Development

Looking at the distribution of the documents from the publications shows that there is a good amount of text supporting health data management and its development technology and policy factors. In 19 publications discussing data management, the concepts of blockchain, healthcare, technology, information, security, network, smart, applications, access, and privacy have been the top 10 discussed. The coexistence of information and healthcare amongst the top 10 discussed concepts shows that the publications have also pointed out the importance of healthcare data management as it was also perceived by the Norwegian citizens as one of the most significant indicators of their country (IMD, 2021). Moreover, the coexistence of smart, technology, blockchain, and applications amongst the top 10 discussed concepts shows how publications relate to the applications of blockchain in the smart asset development framework that defines technology and policy as the driving factors of health data management development (Yigitcanlar et al., 2018). In addition, the coexistence of security, privacy, and access among the top 10 discussed concepts shows how publications prioritize policy-related matters in health data management as it also has been significantly important for both smart asset development and for Scandinavian users (IMD, 2021; Yigitcanlar et al., 2018).

71 e-Health and Blockchain-related publications have discussed the concepts of technology, smart, security, network, information, access, application, management, privacy, and transactions the most. Here also, the coexistence of smart, technology, information, application, and management delivers the importance of blockchain uses in healthcare data management developments based on the smart asset development framework (Yigitcanlar et al., 2018). Technological advancements have pushed healthcare data management toward blockchain; however, security, privacy, and access concerns are also coming along with it (Gatteschi et al., 2018; Ismail et al., 2020). Interestingly, here among the top 10 discussed concepts in e-Health and blockchain, there are security, privacy, and access that confirm their significance in a blockchain-based healthcare data management development.

Finally, 52 security, privacy, GDPR, and e-Health Directorate-related publications have discussed blockchain, data, healthcare, smart, technology, information, security, network, access, and applications as the top 10 concepts. Here also, the coexistence of the mentioned concepts in the publications shows how important and related policy is for the blockchain-based health data management developments (Yigitcanlar et al., 2018). Also, the coexistence of blockchain, security, and access confirms that prominent features of blockchain, like ledgers, decentralization, cryptography, consensuses, and smart contracts are able to establish trust, authorize access, enhance security, and improve transparency (Huumo et al., 2016; Yaga et al., 2018).

5.2 Patient-Centricity of Health Data Management System

Looking at the distribution of the tweets from the users indicates that users are discussing security, privacy, and GDPR-related concepts the most. Only 1% of the tweets are discussing data management and the top 10 discussed concepts are services, utilities, systems, customer, help, analytics, new, time, needs, and research. The coexistence of the mentioned concepts indicates the fact that users prioritize services and utilization of health data management systems the most as it has also been identified by the Directorate for e-Health. Directorate for e-Health defines patients as the most important part of the architecture and entitles them the right to follow their treatments, decide about their treatments, repeat advice during the consultation, seek reassessment, choose selective accession for anyone other than themselves, and know the holders of their information (Directorate for eHealth, 2021). Moreover, the coexistence of the concepts of customers, help, and needs clarifies the importance of user-centricity or patient-centricity in a Scandinavian health data management system as it has also been pointed out that Nordic countries are well-known digital service providers who engage the citizens with public institutions by collecting combinable data from them (Choroszewicz & Alastalo, 2021).

About 28.2% of the discussion of the users have been around e-Health and Blockchain-related concepts. The top 10 discussed concepts are crypto, Bitcoin, NFT, Ethereum, web, Metaverse, decentralized, news, gaming, and project. The coexistence of the mentioned concepts reveals the fact that the users are mostly knowledgeable or concerned about the financial or non-health applications of blockchain; however, blockchain technology savviness has the potential to move the Scandinavian governments towards a creative patient-centered health data management system (Laurini, 2020). For this purpose, patients should become more acquainted with the benefits and health applications of blockchain to accept self-managed blockchain-based health data management. Otherwise, traditional systems and miscomprehension of blockchain's quality hinder the benefit perception and intention of using such systems (El-Gazzar & Stendal, 2020; Shrestha & Vassileva, 2019; Yaga et al., 2018). Norway's capital had been the 3rd smartest city in the world in 2021, and knowledge-intensivity is one of the characteristics of such cities (Laurini, 2020). Accordingly, knowledge can be added as another important development driving factor which is a common good that impacts both the disruptive implementations and users' acceptance (Laurini, 2020). As a result, public and organizational knowledge is an important factor influencing implementation of blockchain-based health data management systems (Laurini, 2020).

Finally, about 70.8% of the discussion of the users has been around security, privacy, and GDPR. It shows that security, privacy, and GDPR-related matters, or in other words, policies, are the most important development driving factor from the point of view of the users. The top 10 discussed concepts are data, new, people, cybersecurity, need, right, GDPR, threat, social, and think. The coexistence of data, right, and GDPR among the top 10 concepts reveals that users are concerned about their rights regarding their data and GDPR is an important righteousness indicator. In a data management context, it delivers the fact that Scandinavian users are aware of the global movement of the governments towards promoting the rightfulness of data subjects in making proactive decisions regarding sharing and transferring their data (Choi et al., 2021). This global movement pushes the health sector and health data management systems to become patient-centered (Choi et al., 2021). In addition, the coexistence of cybersecurity and threat among the top 10 discussed concepts confirms the fact that Scandinavian users signify the security of personal data used in public digital services, like health data management services (IMD, 2021). They are worried about threats to their data; however, blockchain can promote the security and reliability of patients' data.

The combination of blockchain and the transformation of healthcare to a patient-centered approach helps consolidate and exchange patient data securely across interoperable health systems (Hölbl et al., 2018). Consent-based use of personal health data is vital and requires a specific personalized service architecture for patient-centered health data management (Choi et al., 2021). A patient-centered health data management system elevates personalized treatments, and entitles the rights of control, access, and sharing of data to the patients; however, a shift from the current centralized to a decentralized health care system is required (Jabarulla & Lee, 2021). Blockchain, distributed ledgers, smart contracts, consensuses, etc. help overcome the challenges of patient-centricity, security, and privacy (Shi et al., 2020). However, it brings up challenges like cost, data size, and privacy (Alamri et al., 2021).

5.3 Attitudes

Interestingly, all the publications and users have shown positive sentiments about the health data management blockchain-based development. Publications discussing data management have shown positive IG sentiment of 0.10 and users discussing data management have shown positive IG sentiment of 0.17. However, these are not clearly positive attitudes, and based on the most discussed concepts, these may have been impacted by security, privacy, and access concerns from the publications' side, and by the help and needs concerns from the users' side. It is also indicated by the scholars that policy-based accessing is one of three fundamental processes of any data management system (Hu et al., 2021; Mondal et al., 2019). Although Scandinavian users are not clearly positive about data management systems, Nordic countries are increasingly using analytical tools like health data management systems to provide management with information about clinical and financial aspects of the organization (Choroszewicz & Alastalo, 2021). Accordingly, a patient-centered approach that perceives the patients as the owner of their data can increase engagement and interest in patients for managing their health data (Choroszewicz & Alastalo, 2021; Gil-Garcia, as cited in Ismagilova et al., 2019; Jabarulla & Lee, 2021).

When discussing e-Health and Blockchain, publications and users have shown a positive IG sentiment of, respectively, 0.11 and 0.09. Here also there is positivity, but not a clear positivity. As a result, from the publications' side, there are concerns about security and privacy. However, blockchain-based medical data protection is a powerful tool that provides confidentiality, authentication, integrity, and defined access if designed based on a framework that signifies architecture, data integrity, data sharing, access control, distributed data, patient encryption key, framework, and algorithm (Bodkhe et al., 2020; Zhang et al., 2020). Positivity from the users' side is not valid enough in our context, because they have been mostly discussing financial aspects in their discussions, but they have shown a positive attitude toward the uses and applications of blockchain.

Finally, when discussing security, privacy, GDPR, and regulations of the Directorate for e-Health, publications and users are showing a positive IG sentiment of, respectively, 0.13 and 0.14. Here also, the positivity is not highly clear that may have been impacted by privacy, security, and access concerns from the publications side, and by GDPR, need, right, cybersecurity, and threat concepts from the users' side. Although the role of the data controller and the right of erasure defined by the GDPR act are two problematic requirements of any blockchain-based health information system, solutions like off-chain storage and sub-chain storage have been introduced that help the logical deletion and real deletion of data (Haque et al., 2021; Hasselgren et al., 2020; Kuperberg, 2020). Also, controller role issues are addressable by defining different role-based nodes on blockchains. Like administrator, controller, contractor, etc. nodes (Haque et al., 2021). Moreover, a usable blockchain-based system empowers the users with unlosable data ownership and incentivized data sharing (Shrestha & Vassileva, 2019).

5.4 Relations

When publications discuss the elements of health data management development framework, their notions indicate remarkable internal correlation, as illustrated before in Figure 18. An internal correlation of 0.98 to 0.99 shows the accuracy of the notion exploration around blockchain-based health data management development. On the other hand, when users discuss either data management, e-Health/Blockchain, or Security/Privacy/GDPR concepts, their notions have a real low internal correlation ranging from 0.12 to 0.19. This variation of notion indicates a rich unbiased data collection from anonymized Scandinavian users on Twitter.

After all, since this thesis is considering patient-centered developments, the correlations will be explained from their point of view. As illustrated in Figure 18, 50% of the relationships are merely changing with a minimum internal correlation of 0.98. Instead, when users discuss either data management, e-Health/Blockchain, or Security/Privacy/GDPR concepts, their notions' relation with these fixed factors changes meaningfully. In this context, data management is the smart asset, Blockchain is the technology intertwined with e-Health, and GDPR is the policy factor. For instance, when users discuss data management, their notions correlate with policy and technology factors, respectively, 0.11 and 0.55., and correlate with data management concepts in the publications 0.37. It indicates that when they discuss data management, they are almost publicly aware of the importance of technology factors. Nonetheless, they are not publicly aware of security, privacy, GDPR, or Directorate for e-Health-related matters.

In contrast, when users discuss Blockchain and e-Health, their notions correlate 0.55 with the publications' technology factors, and 0.54 with the data management and policy factors. It indicates that when users discuss Blockchain and e-Health, they are almost publicly aware of the importance of data management and its development factors, technology, and policy factors. Nonetheless, when users discuss Security/Privacy/GDPR concepts, their notions show a non-significance correlation of 0.11 with all the other elements. It indicates that, when users discuss Security/Privacy/GDPR, they are not aware of the important applications and features of blockchain in health data management. They also are not publicly aware of the importance of health data management and its GDPR-related matters. This last result may have been impacted by the ongoing war between Russia and Ukraine. Invalid records were removed, but there is uncertainty to some extent. To conclude, a Scandinavian blockchain-based patient-centered health data management is highly dependent on the knowledge of the citizens. Not only technological knowledge but also regulative data protection knowledge. In this respect, this thesis proposes a conceptual four-factor development model for health data management in Scandinavia. This framework includes public knowledge as the fourth element.

5.5 Proposed Development Solution

Scandinavian governments are increasingly digitalizing public services like health data management, and Norway, as a sample, has been doing it leadingly in the framework of smart cities (Choi et al., 2021; Choroszewicz & Alastalo, 2021; IMD, 2021). If a blockchain-based patient-centered development is ever required, there should be a shared common knowledge factor for successful implementation (Laurini, 2020). This common knowledge factor includes public knowledge regarding the regulative, technological, and core factors of health data management, as illustrated in Figure 19. It has been previously mentioned as technology savviness, one of the smartness dimensions in the government (Gil-Garcia et al., 2016). However, in the case of health data management development, it is a driving factor that should be considered an infrastructure for the development (Laurini, 2020).

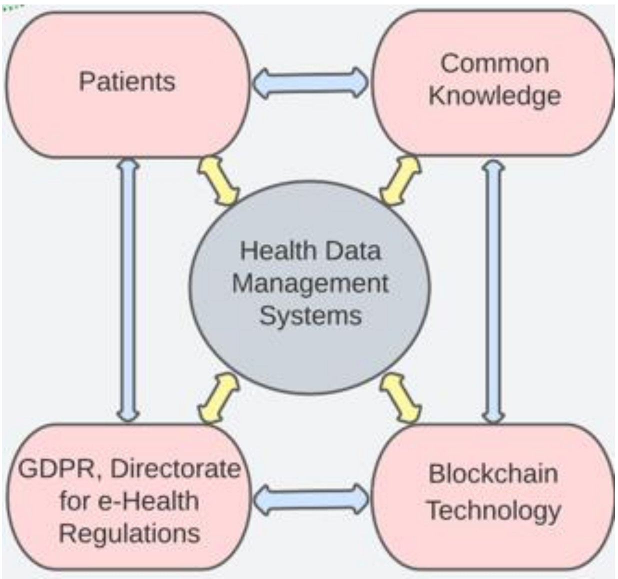

Patients
Common Knowledge
Health Data Management Systems
GDPR, Directorate for e-Health Regulations
Blockchain Technology

Figure 19 - Proposed Blockchain-Based Patient-Centered Health Data Management Development Framework

In the literature, complexity, miscomprehension, and user acceptance were mentioned as some of the challenges of blockchain technology (El-Gazzar & Stendal, 2020; Yaga et al., 2018). However, the proposed common knowledge factor aims to fill the gap and notify the users of health data management systems about policy rights, interoperable abilities, technological utilities, and self-managed health data protection.

This framework proposes the concepts of Electronic Health Wallet (EHW), off-chain permissioned blockchain, and tree-based ledger and linear sub-chains to create a health data management system that stores data lawful, fair, transparent, restricted, encrypted, accurate, erasable, integrated, confident, and accountable, as requested by GDPR and the Directorate for e-Health (Directorate for eHealth, 2021; EU, 2022; Vigot & Bussche, 2017). This EHW allows the patients to access their up-to-date health data in a real-time manner (Alamri et al., 2021). EHW should be able to collect data from IoT devices, EHRs, and patients, and use the measures provided by the data management layer to authorize users, manage accessions, and apply incentives (Alamri et al., 2021). This thesis proposes to minimize the main actors of the EHWs to patients and healthcare service providers.

Proposed off-chain permissioned blockchain can restrict reading and writing access, authorize roles, establish trust, and interoperate effectively on systems (Hylock & Zheng, 2019; Yaga et al., 2018). Or instead, the proposed tree-based ledger and linear sub-chains can store data online and provide legal deletion for data along with all other benefits (Haque et al., 2021; Hasselgren et al., 2020; Kuperberg, 2020). In the literature, data erasure and data controller role were mentioned as two problematic GDPR requirements for blockchains (Esmel et al., 2021; Haque et al., 2021; Kuperberg, 2020). However, the proposed health data management system seems to overcome these challenges.

For the common knowledge concept, this thesis proposes a federated public blockchain that requires the participants to authorize themselves and get verified through the PBFT consensus algorithm. The Practical Byzantine Fault Tolerance (PBFT) consensus includes five phases of request, pre-prepare, prepare, commit, and reply. The Practical Byzantine Fault Tolerance (PBFT) is a less complex and highly practical consensus including five phases of request, pre-prepare, prepare, commit, and reply. PBFT uses three nodes for previously mentioned phases to reach a consistent reply for any request (Zhang & Lee, 2020).

This way, PBFT in a federated blockchain can help authorize patients, regulative bodies, data controllers, and third parties (Dinh et al., 2018; Yaga et al., 2018). All parties go through the PBFT algorithm to authorize themselves as a propagative block, and then each governing block uses localized smart contracts to provide peer-to-peer interaction with other local governing blocks (Huumo et al., 2016; Yaga et al., 2018). In the literature, security attacks, explicit consent, connection with real-world, resource usage, scalability, and regulatory uncertainties were mentioned as the other challenges of blockchain technology (Chen et al., 2019; Esmel et al., 2021; Huumo et al., 2016; Truong, et al., 2019; Yaga et al., 2018; Yaqoob et al., 2021). However, integration of Ripple, which is a PBFT-based blockchain, and federated blockchains can address these issues by acting absolute with 20% fault tolerance, negligible power consumption, good scalability, governing or regulative capabilities, consent management (Haque et al., 2021; Jabarulla & Lee, 2021; Majdoubi et al., 2021; Zhang & Lee, 2020).

To summarize the development model, it is a three-layered blockchain-based patient-centered health data management system. In the literature, patients' participation, privacy, public insight, and governing bodies were the main concerns (Ismail et al., 2020). However, different three-layered solutions tackle these challenges by including a decentralized permissioned blockchain-based data store layer, a governing federated public blockchain layer, and an AI layer (Jabarulla & Lee, 2021; Saha et al., 2021). The proposed solution has adopted the goal architecture of the Directorate for e-Health to customize these three-layered solutions to meet the common sharing platform, identification, joint API, authorization, consent management, logging, and common health information requirements of the Directorate for e-Health (Directorate for eHealth, 2021).

The bottom layer uses a decentralized blockchain-based data store layer with only two actors who are patients and digital health providers. This way, the proposed solution will also comply with the goal architecture of the Directorate for e-Health. Patient Records Act 18, Patient and User Rights Act 5-1, Privacy Ordinance Article 15, and Health Personnel Act 41 are addressed in this relation (Directorate for eHealth, 2021).

The middle layer is a federated blockchain layer that includes loosely coupled blockchains within a common knowledge main blockchain. The directorate for e-Health is the data controller sub-blockchain, patients are the private EHWs sub-blockchain, digital health providers are the EHRs sub-blockchain, and the third parties are the GDPR sub blockchain. In fact, the GDPR sub-blockchain consists of any third party that requests to process data. They should first get verified in that sub-blockchain to approve GDPR-compliancy. All these sub-blockchains interact based on FHIR and HL7 universal health data exchange frameworks. Trusted APIs help translatable interoperable data sharing in a common platform (Jabarulla & Lee, 2021; Saha et al., 2021; Yaga et al., 2018).

Finally, the top layer uses machine learning and deep learning to analyze the health data from the other layers and enhance public insight and decision-making (Ismail et al., 2020; Jabarulla & Lee, 2021). This way, all the requirements of the goal architecture of the Directorate for e-Health are addressed in a GDPR-compliant way. Figure 20 illustrates the proposed architecture for a patient-centered blockchain-based health data management system from a Scandinavian point of view.

Blockchains include 6 working layers which are data, network, consensus, contract, service, and application (Yang, 2019).

The proposed architecture considers the blockchain-based health data management transactions introduced in the literature, as illustrated in Figure 7 (Bodkhe et al., 2020). Blockchain applications can be compared to each other based on their architecture, data integrity, data sharing, access control, distributed data, patient encryption key, framework, and algorithm (Bodkhe et al., 2020). In this regard, decentralized data is stored integrated, patients' EHRs are not accessed without their individual encryption keys, common sharing platforms provide universal sharing, smart contracts handle privacy, and sub-blockchains handle local governing.

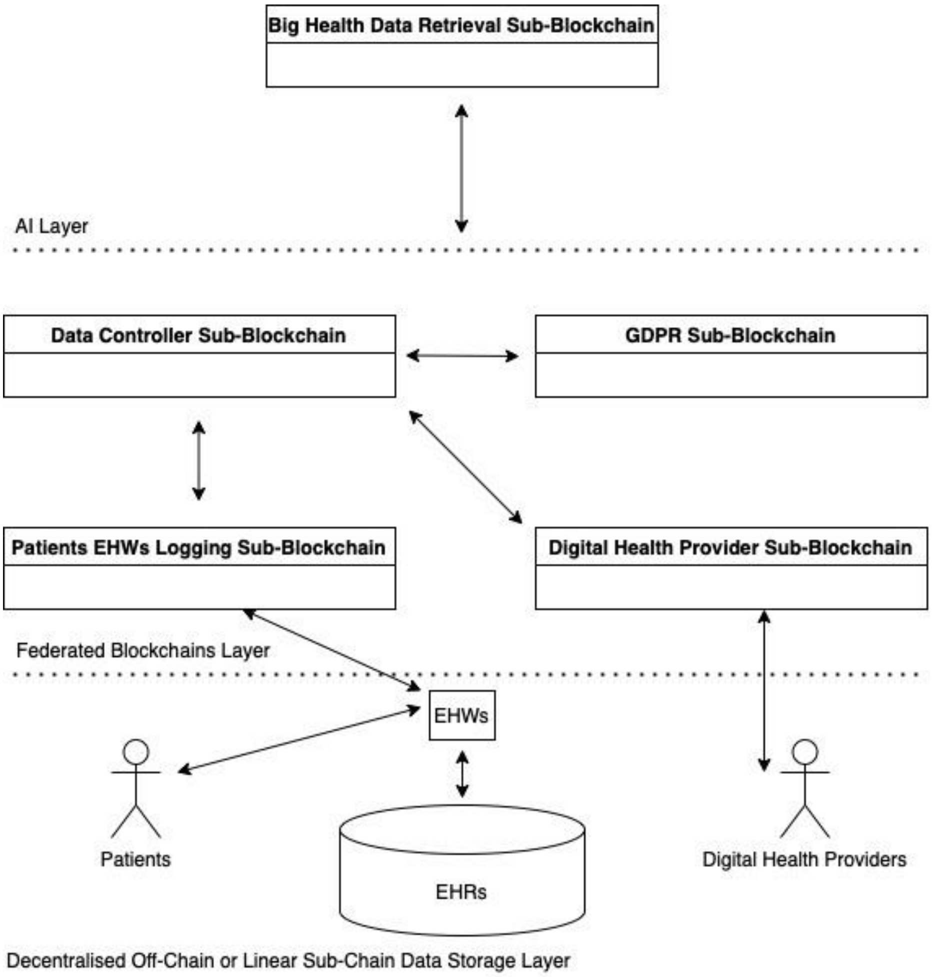

Big Health Data Retrieval Sub-Blockchain
AI Layer
Data Controller Sub-Blockchain
GDPR Sub-Blockchain
Patients EHWs Logging Sub-Blockchain
Digital Health Provider Sub-Blockchain
Federated Blockchains Layer
EHWs
Patients
EHRs
Digital Health Providers
Decentralised Off-Chain or Linear Sub-Chain Data Storage Layer

Figure 20 - Conceptual Blockchain-Based Patient-Centered Health Data Management System from A Scandinavian Point of View

As it is depicted, every stakeholder should authorize and identify themselves to the data controller and no one has access to the EHRs except the ones who request accession through EHWs and smart contracts.

6 Implications and Future Work

Based on the results from publications and Twitter users, this thesis has pointed out the necessity of common technological and regulative knowledge for health data management development in Scandinavia. As a result, this thesis considers the common knowledge factor as the fourth crucial element that drives health data management development. Furthermore, this thesis proposes a blockchain-based solution for the development of health data management in Scandinavia which fills the patient-centricity and GDPR compliancy gap. Accordingly, proposed health data management development framework has the potential to be of use for Scandinavian e-Health providers and regulative bodies who consider any type of health data management development. In addition, the blockchain-based solution has the potential to be tested in a small voluntary scale for citizens who are willing to self-manage their health data. Nonetheless, in any case, all the stakeholders are required to reach a common knowledge which is defined and controlled by the Directorate for e-Health.

This thesis clarified that for a patient-centered approach, in the first step, responsible organizations must raise public awareness regarding privacy rights, technology, and regulations. Next, governments should update their citizens about driving factors of health data management and concurrently provide the infrastructures for possible future developments to become more agile and smart. In this case, although blockchain is immature, it is anticipated to be a powerful tool. As a result, since blockchain-based solution is a futuristic topic, it is better that the governments who consider it start the process of public knowledge creation years before the implementation.

In addition, further research is required to focus on richer data collection, public knowledge creation, and infrastructural readiness for successful implementation. This thesis includes a rich data set that still has many untouched or uncovered dimensions that may be helpful for other fields of study that are intertwined with data management, security, privacy, blockchain, Twitter, or e-health. Further research can consist of exploration of exact Scandinavian-based health data transaction, public acceptance of self-data management, infrastructural readiness, or other meaningful data pattern extraction from the rich Twitter collected data set with 90 columns of more than 17000 rows of tweets.

7 Conclusion

On one hand, digitalization has empowered healthcare with various digital health services like electronic health records, patient monitoring, disease diagnostic, treatment enhancement, etc. On the other hand, it has brought big data, reliability, privacy, and security challenges and has made healthcare highly dependent on health data management (IMD, 2021; Mondal et al., 2019). Moreover, recently, patient-centricity in health data management has been emerging as a must (Choi et al., 2021). Different regulations like the GDPR act, HIPAA act, or the regulations of the Directorate for e-Health all share the same rights which define the patients as the owner of their health data and the ones who make proactive decisions regarding sharing their data (Ahram et al., 2017; EU, 2022; Vigot & Bussche, 2017).

In addition, Nordic countries are prominent world leaders in public service digitalization with the help of data they collect from their citizens. This elevates the citizens' engagement with the public institutions (Choroszewicz & Alastalo, 2021). One of these public services being digitalized in

Scandinavia increasingly is the health services (Choroszewicz & Alastalo, 2021). Norway, as a sample of Scandinavia, has been digitizing its services in the framework of smart cities (IMD, 2021). As a result, a patient-centered health data management system is crucial in a Scandinavian smart city context to sustain smartness (IMD, 2021; Laurini, 2020). Accordingly, this thesis has investigated health data management development in the framework of smart asset development (Yigitcanlar et al., 2018). In this regard, health data management development driving factors are explored both from the point of view of the publications and Scandinavian Twitter users. Also, since Norway, as the main focus of this thesis, has a rich high-tech infrastructure, this thesis explores possible blockchain-based solutions for the development of current health data management in Scandinavia (IMD, 2021). It is believed that blockchain can enhance security and reliability; however, it may bring some challenges also (Hölbl et al., 2018).

To find out what is the common belief around blockchain-based patient-centered health data management systems, the concepts of health data management systems, blockchain as the technological development driving factor, and GDPR and regulations of the Directorate for e-Health as the regulative development driving factors were explored thoroughly in a systematic literature review. In this regard, core concepts, early designs, recent improvements, and blockchain-based solutions were investigated. Also, challenges and opportunities were discussed, and prominent designs were explained in detail to find what blockchain capabilities can resolve related issues. Furthermore, to acquire a holistic approach toward patient-centricity, more than 17000 tweets regarding data management, e-health and blockchain, and security, privacy, and GDPR were mined. After a thorough literature review, the RStudio Programming tool was used to extract the most discussed concepts, emotions, beliefs, and relations from publications and tweets.

Results from the literature review revealed that blockchain, GDPR, and regulations of the Directorate for e-Health have the potential to create an integrated interoperable patient-centered health data management system; however, there are challenges like erasure right, data controller role, scalability, miscomprehension, power consumption, data size, security, user acceptance, self-data managing, immaturity, etc. (El-Gazzar & Stendal, 2020; Esmel et al., 2021; Haque et al., 2021; Shrestha & Vassileva, 2019; Yaga et al., 2018). Most of these challenges are addressable by federated blockchains, sub-chains, off-chain storage, PBFT consensus, smart contracts, etc. (Alamri et al., 2021; Esmel et al., 2021; Haque et al., 2021; Hasselgren et al., 2020; Hylock & Zheng, 2019; Kuperberg, 2020; Yaga et al., 2018). Nonetheless, some challenges like miscomprehension or blockchain immaturity are yet required to be addressed.

In addition, results from text mining in publications and tweets revealed that when publications and users discuss blockchain and e-Health, their notions have a higher correlation with each other and with other development driving factors being explored. Instead, when they discuss security, privacy, and GDPR, their notions do not correlate at all with each other or with other development driving factors. Accordingly, this thesis found out that when users have more technological knowledge, they are much more aware of their privacy and policy rights. However, when they discuss policy-related matters, they are not aware of the capabilities of technology in tackling their issues. Also, when they discuss data management, they are aware of the technological and policy factors. Eventually, these findings reveal that if a health data management development is ever required, a common knowledge that is shared between the stakeholders is crucial. This common knowledge should encompass roles, rights, technological capabilities, and regulative bodies. As a result, this thesis proposed a fourth element to be added to the smart asset development driving factors, the common knowledge factor. Sentimental results from the publication and tweets mining revealed that both publications and users are thinking positively about data management, e-health and blockchain, and security, privacy, and GDPR. Eventually, the goal architecture defined by the Directorate for e-Health was explored in detail and a conceptual blockchain-based solution for patient-centered health data management development in Scandinavia was proposed.

Taking the personal data concerns and high density of digitalization in Norway into account reveals that becoming one of the smartest countries in the world has costs that together with its benefits work like a double-edged sword. On one hand, rich infrastructure and high technological projections, and on the other hand, the urgency of health, security, and personal data throughout the digitalization process (IMD, 2021). This thesis has proposed a customized health data management development framework and a conceptual blockchain-based patient-centered health data management system from a Scandinavian point of view to help Scandinavian digital health providers and regulative bodies understand the driving factors for such developments and consider blockchain-based solutions in the future. As pointed out previously, the common knowledge factor is a key to successful implementation, and it takes years to build it.

Accordingly, if Scandinavian digital health providers and regulative bodies consider any type of health data management development, it is better to use the proposed development framework and consider creating common knowledge regarding the rights, technology, and self-managed health data well in advance. If the proposed conceptual blockchain-based solution is considered, the common knowledge factor can include the capabilities and utilities of blockchain technology, the

GDPR, HIPAA, and Directorate for e-Health regulations, and the roles of the patient, data controller, digital health provider, and third parties.

References

Aggarwal, S. & Kumar, N. (2020). Chapter Twenty-Four – Healthcare System. *Advances in Computers, 121,* 483-493. https://doi.org/10.1016/bs.adcom.2020.08.024

Ahram, T., Sargolzaei, A., Sargolzaei, S., Daniels, J. & Amaba, B. (2017). *Blockchain Technology Innovations* [Conference Presentation]. IEEE Technology & Engineering Management Conference (TEMSCON), Santa Clara. https://doi.org/10.1109/TEMSCON.2017.7998367

Alamri, B., Javed, I. T. & Margaria, T. (2021). *A GDPR-Compliant Framework for IoT-Based Personal Health Records Using Blockchain* [Conference Presentation]. 11th IFIP International Conference on New Technologies, Mobility and Security (NTMS), Paris. https://doi.org/10.1109/NTMS49979.2021.9432661

Al-asmari, A. M., Aloufi, R. I. & Alotaibi, Y. (2021). A Review of Concepts, Advantages and Pitfalls of Healthcare Applications in Blockchain Technology. *International Journal of Computer Science and Network Security, 21*(5), 199-210. https://doi.org/10.22937/IJCSNS.2021.21.5.28

Bodkhe, U., Tanwar, S., Parekh, K., Khanpara, P., Tyagi, S. & Kumar, N. (2020). Blockchain for Industry 4.0: A Comprehensive Review. *IEEE Access, 8,* 79764-979800. https://doi.org/10.1109/ACCESS.2020.2988579

Bryman, A. & Bell, E. (2011). *Business Research Methods*. (3rd ed.). Oxford University Press.

Chen, J., Lv, Z. & Song, H. (2019). Design of Personnel Big Data Management System Based on Blockchain. *Future Generation Computer Systems, 101,* 1122-1129. https://doi.org/10.1016/j.future.2019.07.037

Chkirbene, Z., Mohamed, A., Erbad, A. & Guizani, M. (2020). *Smart Edge Healthcare Data Sharing System* [Conference Presentation]. International Wireless Communications and Mobile Computing (IWCMC), Limassol. https://doi.org/10.1109/IWCMC48107.2020.9148195

Choi, W., Chun, J. W., Lee, S. J., Chang, S. H., Kim, D. J. & Choi, I. Y. (2021). Development of a MyData Platform Based on The Personal Health record Data Sharing System in Korea. *Applied Sciences, 11*(17). https://doi.org/10.3390/app11178208

Choroszewicz, M. & Alastalo, M. (2021). Organisational and professional hierarchies in a data management system: public–private collaborative building of public healthcare and social services in Finland. *Information, Communication & Society.* https://doi.org/10.1080/1369118X.2021.1942952

Dinh, T. T. A., Liu, R., Zhang, M., Chen, G., Ooi, B. C. & Wang, J. (2019). Untangling Blockchain: A Data Processing View of Blockchain Systems. *IEEE Transactions on Knowledge and Data Engineering, 30*(7), 1366-1385. https://doi.org/10.1109/TKDE.2017.2781227

Directorate for eHealth. (2021). *Goal Architecture for Data Sharing in Healthcare Sector.* (HITR 1231:2021). Retrieved from https://www.ehelse.no/standardisering/standarder/malarkitektur-for-datadeling-i-helse-og-omsorgssektoren/_/attachment/inline/a5a908cd-5054-4d21-8eaf-8b795dcb25ea:761793d5dd6b6a1f2b9dd334a44e3a754d5b88e6/Målarkitektur%20for%20datadeling%20i%20helse-%20og%20omsorgssektoren%20(HITR%201231_2021).pdf

El-Gazzar, R. & Stendal, K. (2020). Blockchain in Health Care: Hope or Hype?. *Journal of Medical Internet Research, 22*(7). https://doi.org/10.2196/17199

Esmel, C. D., Roca, J. C., Viejo, A. & Ferrer, J. D. (2021). *Lightweight Blockchain-Based Platform for GDPR-Compliant Personal Data Management* [Conference Presentation]. IEEE 5th International Conference on Cryptography, Security, and Privacy (CSP), Zhuhai. https://doi.org/10.1109/CSP51677.2021.9357602

European Union. (2022). *What is GDPR, the EU's new data protection law?.* GDPR.EU. https://gdpr.eu/what-is-gdpr/

Foltz, P. W., Kintsch, W. & Landauer, T. K. (1998). The Measurement of Textual Coherence With Latent Semantic Analysis. *Discourse Processes, 25*(2-3), 285-307. https://doi.org/10.1080/01638539809545029

Fu, L., Ding, S. & Chen, T. (2010). *Clinical Data Management Systems* [Conference Presentation]. International Conference on Biomedical Engineering and Computer Science, Wuhan. https://doi.org/10.1109/ICBECS.2010.5462386

Gatteschi, V., Lamberti, F., Demartini, C., Pranteda, C. & Santamaría, V. (2018). To Blockchain or Not to Blockchain: That is the Question. *IT Professionals, 20*(2), 62-74. https://doi.org/10.1109/MITP.2018.021921652

Gil-Garcia, J. R., Zhang, J. & Puron-Cid, G. (2016). Conceptualizing Smartness In Government: An Integrative and Multi-Dimensional View. *Government Information Quarterly, 33*(3), 524-534. https://doi.org/10.1016/j.giq.2016.03.002

Greenes, R. A., Pappalardo, A. N., Marble, C. W. & Barnett, G. O. (1969). Design and Implementation of a Clinical Data Management System. *Computer and Biomedical Research, 2,* 469-485. https://doi.org/10.1016/0010-4809(69)90012-3

Haque, A. B., Islam, A. K. M. N., Hyrynsalmi, S., Naqvi, B. & Smolander, K. (2021). GDPR Compliant Blockchains – A Systematic Literature Review. *IEEE Access, 9,* 50593-50606. https://doi.org/10.1109/ACCESS.2021.3069877

Hasselgren, A., Wan, P. K., Horn, M., Kralevska, K. Gligoroski, D. & Faxvaag, A. (2020). GDPR Compliance for Blockchain Applications in Healthcare. *Arxiv Cornell University.* https://doi.org/10.48550/arXiv.2009.12913

Hu, H., Qi, F., Zhang, H., Tian, H. & Luo, Q. (2021). The Design of A Data Management System at HEPS. *Journal of Synchrotron Radiation, 28,* 169-175. https://doi.org/10.1107/S1600577520015167

Huumo, Y. J., Ko, D., Choi, S., Park, S. & Smolander, K. (2016). What is Current Research on Blockchain Technology? – A Systematic Review. *PLoS ONE, 11*(10). https://doi.org/10.1371/journal.pone.0163477

Hylock, R. H. & Zeng, X. (2019). A Blockchain Framework for Patient-Centered Health Records and Exchange (HealthChain): Evaluation and Proof-of-Concept Study. *Journal of Medical Internet Research, 21*(8). https://doi.org/10.2196/13592

Hölbl, M., Kompara, M., Kamišalić , A. & Zlatolas, L. N. (2018). A Systematic Review of the Use of Blockchain in Healthcare. *Symmetry, 10*(10). https://doi.org/10.3390/sym10100470

International Institute for Management Development. (2021). *Smart City Index 2021*. (3). Retrieved from https://www.imd.org/link/d087d9f47fc84dd7bcab146e5c9601dc.aspx

Ismagilova, E., Hughes, L., Dwivedi, Y.K., & Raman, K.R. (2019). Smart cities: Advances in research— An information systems perspective. *International Journal of Information Management, 47.* 88-100. https://doi.org/10.1016/j.ijinfomgt.2019.01.004

Ismail, L. Materwala, H. & Hennebelle, A. (2021). A Scoping Review of Integrated Blockchain-Cloud (BcC) Architecture for Healthcare: Applications, Challenges, and Solutions. *Sensors, 21*(11). https://doi.org/10.3390/s21113753

Ismail, L., Materwala, H., Karduck, A. P. & Adem, A. (2020). Requirements of Health Data management Systems for Biomedical Care and Research: Scoping Review. *Journal of Medical Internet Research, 22*(7). https://doi.org/10.2196/17508

Ismail, T., Touati, H., Hajlaoui, N. & Hamdi, H. (2020). *Hybrid and Secure E-Health Data Sharing Architecture in Multi-Clouds Environment* [Conference Presentation]. International Conference on Smart Homes and Health Telematics (ICOST), Tunisia. https://doi.org/10.1007/978-3-030-51517-1_21

Jabarulla, M. Y. & Lee, H. N. (2021). A Blockchain and Artificial Intelligence-Based, Patient-Centric Healthcare System for Combatting the COVID-19 Pandemic: Opportunities and Applications. *Healthcare, 9*(8). https://doi.org/10.3390/healthcare9081019

Kanakaraj, M. Mohana, R. & Guddeti, R. (2015). *Performance Analysis of Ensembling Methods on Twitter Sentiment Analysis Using NLP Techniques* [Conference Presentation]. IEEE 9th International Conference on Semantic Computing (IEEE ICSC), Anaheim. https://doi.org/10.1109/ICOSC.2015.7050801

Kennedy, O., Ahmad, Y. & Koch, C. (2011). *DBToaster: Agile Views in a Dynamic Data Management System* [Conference Presentation]. CIDR 2011, Fifth Biennial Conference on Innovative Data Systems Research, Asilomar. https://infoscience.epfl.ch/record/165840/files/CIDR11_Paper38.pdf

Khatri, S., Alzahrani, F. A., Ansari, M. T. J., Agrawal, A., Kumar, R. & Khan, R. A. (2021). A Systematic Analysis on Blockchain Integration With healthcare Domain: Scope and Challenges. *IEEE Access, 9,* 84666-84687. https://doi.org/10.1109/ACCESS.2021.3087608

Kuperberg, M. (2020). *Towards Enabling Deletion in Append-Only Blockchains to Support Data Growth Management and GDPR Compliance* [Conference Presentation]. IEEE International Conference on Blockchain, Rhodes. https://doi.org/10.1109/Blockchain50366.2020.00057

Laurini, R. (2020). A primer of knowledge management for smart city governance. *Land Use Policy.* https://doi.org/10.1016/j.landusepol.2020.104832

Majdoubi, D., Bakkali, H. E. & Sadki, S. (2021). SmartMedChain: A Blockchain-Based Privacy-Perserving Smart Healthcare Framework. *Journal of Healthcare Engineering, 2021.* https://doi.org/10.1155/2021/4145512

McGhin, T., Choo, K. K. R., Liu, C. Z. & He, D. (2019). Blockchain in Healthcare Applications: Research Challenges and Opportunities. *Journal of Network and Computer Science, 135,* 62-75. https://doi.org/10.1016/j.jnca.2019.02.027

Meng, Y., Huang, Z., Shen, G. & Ke, C. (2019). SDN-Based Security Enforcement Framework for Data Sharing Systems of Smart Healthcare. *IEEE Transactions on Network and Service Management, 17*(1), 308-318. https://doi.org/10.1109/TNSM.2019.2941214

Minckler, T. M., Ausman, R. K., Graham, T., Newell, G. R. & Levine, P. H. (1967). HUMARIS – An Automated Medical Data Management System. *Methods of Information in Medicine, 6*(2), 65-69. https://doi.org/10.1055/s-0038-1636252

Mondal, A. S., Neogy, S., Mukherjee, N. & Chattopadhyay, S. (2019). A Survey of Issues and Solutions of Health Data Management Systems. *Innovations in Systems and Software Engineering, 15,* 155-166. https://doi.org/10.1007/s11334-019-00336-4

Moumen, A. & Mejjad, N. (2021). *Graduates Employability: An Exploratory Literature Review* [Conference Presentation]. The 3rd International Conference on Quantitative and Qualitative Methods for Social Sciences (QQR'21), Kenitra. https://doi.org/10.1051/shsconf/202111905010

Nam, Y., Shin, E., Lee, S., Jung, S., Bae, Y. & Kim, J. (2020). *Global-scale GDPR Compliant Data Sharing System* [Conference Presentation]. International Conference on Electronics, Information, and Communication (ICEIC), Barcelona. https://doi.org/10.1109/ICEIC49074.2020.9051171

Pramanik, M. I., Lau, R. Y. K., Demirkan, H. & Azad, M. A. K. (2017). Smart Health: Big Data Enabled Health Paradigm Within Smart Cities. *Expert Systems with Applications, 87,* 370-383. https://doi.org/10.1016/j.eswa.2017.06.027

Rajawat, A. S., Bedi, P., Goyal, S. B., Shaw, R. N., Ghosh, A. & Aggarwal, S. (2022). AI and Blockchain for Healthcare Data Security in Smart Cities. In Piuri, V., Shaw, R. N., Ghosh, A. & Islam, R. (Ed), *AI and IoT for Smart City Applications* (185-198). Springer. https://doi.org/10.1007/978-981-16-7498-3_12

Rani, V. V. & Rani, K. S. (2016). Twitter Streaming and Analysis Through R. *Indian Journal of Science and Technology, 9*(45), 1-6. https://dx.doi.org/10.17485/ijst/2016/v9i45/97914

RStudio Community. (2022). *Rstudio Community – All Things Rstudio.* Rstudio Community. https://community.rstudio.com

Saha, S., Majumder, A., Bhowmik, T., Basu, A. & Choudhury, A. (2021). *A Healthcare Data Management System on Blockchain Framework* [Conference Presentation]. International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), Pune. https://doi.org/10.1109/SMARTGENCON51891.2021.9645890

Shakil, K. A., Zareen, F. J., Alam, M. & Jabin, S. (2017). BAMHealthcloud: A Biometric Authentication and Data Management System for Healthcare Data in Cloud. *Journal of King Saud University – Computer and Information Science, 32*(1), 57-64. https://doi.org/10.1016/j.jksuci.2017.07.001

Shi, S., He, D., Li, L., Kumar, N., Khan, M. K. & Choo, K. K. R. (2020). Applications of Blockchain in Ensuring the Security and Privacy of Electronic Health Record Systems: A Survey. *Computers and Security, 97.* https://doi.org/10.1016/j.cose.2020.101966

Shrestha, A. K. & Vassileva, J. (2019). *User Acceptance of Usable Blockchain-Based Research Data Sharing System: An Extended TAM-Based Study* [Conference Presentation]. First IEEE International Conference on Trust, Privacy, and Security in Intelligent Systems and Applications (TPS-ISA), Los Angeles. https://doi.org/10.1109/TPS-ISA48467.2019.00033

Vigot, P. & Bussche, A. V. D. (2017). *The EU General Data Protection Regulation (GDPR) – A Practical Guide.* Springer Nature. https://doi.org/10.1007/978-3-319-57959-7

Webster, J. & Watson R. T. (2002). Analyzing The Past To Prepare For The Future: Writing A Literature Review. *MIS Quarterly, 26*(2), xiii-xxiii. http://www.jstor.org/stable/4132319

World Health Organization. (2022). *Human Development Index.* Nutrition Landscape Information System (NLiS). https://www.who.int/data/nutrition/nlis/info/human-development-index

Wu, P., Yi, W. J. & Saniie, J. (2016). *Security Assessment for Personal Health Data Management System* [Conference Presentation]. IEEE International Conference on Electro Information technology (EIT), Grand Forks. https://doi.org/10.1109/EIT.2016.7535277

Yaga, D., Mell, P., Roby, N. & Scarfone, K. (2018). *Blockchain Technology Overview.* (NISTIR8202). Retrieved from https://doi.org/10.6028/NIST.IR.8202

Yang, L. (2019). The Blockchain: State-of-the-Art and Research Challenges. *Journal of Industrial Information Integration, 15,* 80-90. https://doi.org/10.1016/j.jii.2019.04.002

Yaqoob, I., Salah, K., Jayaraman, R. & Al-Hammadi, Y. (2021). Blockchain for Healthcare Data Management: Opportunities, Challenges, and Future Recommendations. *Neural Computing and Applications.* https://doi.org/10.1007/s00521-020-05519-w

Yigitcanlar, T., Kamruzzaman, M., Buys, L., Loppolo, G., Sabatini-Marques, J., Moreira da Costa, E. & Yun, J. J. (2018). Understanding "Smart Cities": Intertwining Development Drivers With Desired Outcomes In A Multidimensional Framework. *Cities, 81,* 145-160. https://doi.org/10.1016/j.cities.2018.04.003

Zhang, J., Zhong, S., Wang, T., Chao, H. C. & Wang, J. (2020). Blockchain-Based Systems and Applications: A Survey. *Journal of Internet Technology, 21*(1), 1-14. https://jit.ndhu.edu.tw/article/view/2217

Zhang, S. & Lee, J. H. (2020). Analysis of the Main Consensus Protocols of Blockchain. *ICT Express, 6*(2), 93-97. https://doi.org/10.1016/j.icte.2019.08.001

List of Tables and Figures