Data Protection and Cybersecurity in Health Information Systems

Nazanin Niayesh

Vienna University of Technology
Bachelor programme Business Informatics
e11809924@student.tuwien.ac.at
Registration No.: 11809924

Supervised by Professor Markus Haslinger

Declaration of Originality

I confirm that the following thesis is original work and was written by me independently and without further assistance.

All information, statements, quotes and figures as well as any other type of content taken from a source are marked in the text and all used sources are cited. No other sources other than the ones listed under references have been used in this work.

Abstract. The increase in cyberattacks on healthcare facilities and data breaches in this branch has exposed the sensitive information of many affected and has put the patients' privacy at risk. To provide a foundation for better understanding the complexity of healthcare cybersecurity issues, this paper discusses the current state of healthcare facilities and provides an overview and comparison of the data privacy regulations in the US and the EU. Some of the deficiencies and areas for improvement in the current technical and legal state of health records are underlined and some recommendations for strengthening the security in health organizations, also in high-pressure situations such as pandemics, are suggested. Furthermore, integrity is introduced as a central concept in data security and the potential of use of (different variations of) the Blockchain technology to ensure the security and integrity of electronic health records is explored.

Table of Contents

Reoccuring Abbreviations (in order of appearance)

HIS	Health Information System
EHR	Electronic Health Record
IoT	Internet of Things
ML	Machine Learning
AI	Artificial Intelligence
HIPAA	Health Insurance Portability and Accountability Act
PPACA/PCA	Patient Protection and Affordable Care Act
DPD	Data Protection Directive
GDPR	General Data Protection Regulation
PDA	Personal Data Act (of Norway)
CERT-RMM	CERT Resilience Management Model
CARE	Concept for Applying Resilience Engineering
DLT	Distributed Ledger Technology
POW	Proof of Work
RPCA	Ripple Protocol Consensus Algorithm
POS	Proof of Stake

1 Introduction

Technological advancements have had numerous benefits for the society and have transformed the everyday life of many people. Also in the healthcare sector, the use of different technologies has become common and will likely only increase in the future due to the improvements these technologies provide. Today, in most healthcare organizations, patient-related data is recorded, updated and generally handled electronically. In the USA for example, Electronic Health Records (EHR) are in use which are real-time records containing a patient's medical and treatment histories, diagnoses, medications and test results amongst others [1]. EHRs are said to improve the quality and efficiency of healthcare services and thus reduce costs significantly [1]. These EHRs and other electronic medical records contain large amounts of medical information about the patients, which may facilitate patient care, but also makes them a great target for cyberattacks.

According to a report by Critical Insight based on statistics published by the US Department of Health and Human Services (HHS), the number of healthcare breaches and individuals affected by them is increasing every year, with an 84% increase in the total number of breaches between 2018 and 2021 alone [2, p. 3]. The total number of affected individuals in the US has more than tripled from 14 million in 2018 to 45 million in 2021 [2, p. 3]. According to CNN, the European Union Agency for Cybersecurity (ENISA) reported a 47% increase in attacks on hospitals and healthcare networks in 2020 where the attackers also targeted the most vital services for the COVID-19 pandemic [3]. The trend of increasing attacks on healthcare organizations and electronic health systems is similar in other parts of the world, which makes the issue of cybersecurity and protection of the patient data pressing and relevant today. Specifically in healthcare, data breaches may lead to patients being blackmailed or cause societal consequences for patients (e.g. workplace discrimination) if released publicly without the patient's knowledge or consent. Successful malicious attacks may also lead to service outage (e.g. in case of Denial-of-Service attacks as described in the following chapters) or otherwise prevent patients from receiving needed treatments or medication which may have life endangering consequences. Recognizing the importance of personal data privacy, Article 12 of the Universal Declaration of Human Rights (UDHR) of the United Nations (UN) states that privacy is a human right and that a person has the right to the protection of the law against attacks or interference with their privacy [4]. The UDHR has been implemented in many countries into regulations and definition of rights. The Charter of Fundamental Rights of the European Union for example also grants a person the right to privacy in its Article 7 and more specifically the right to the protection of personal data in its Article 8 [5].

This work aims to suggest improvements to the current legal and technological state of data protection and cybersecurity in Health Information Systems (HIS). A Health Information System refers to any system designed to collect, process, report and generally manage all healthcare-related data (e.g. EHRs, health facility and community data, surveillance information, supply chain information) to improve the efficiency and effectiveness of health services [6]. In the first following chapter, the current technologies in use in healthcare systems as well as measures used to ensure the security of these technologies are discussed. The current legal situation is illustrated by comparing data privacy regulations in the US and in Europe, specifically the European Union (EU) and European Economic Area (EEA). There is also concrete examples of two electronic HIS, namely that of Austria as an EU country and Norway as a non-EU country to further illustrate possible differences between different HIS within Europe. Additionally, this work identifies two central concepts in data protection, namely resilience and data integrity, which are discussed in the corresponding individual chapters. Enforcing resilience in a healthcare organization, for example through the methods suggested in the corresponding chapter, will further strengthen its existing cybersecurity measures and will not only reduce the possibility of successful attacks, but also facilitate recovery in case of successful attacks and breaches. Ensuring data integrity is especially important regarding health records due to their sensitive information which would have life-threatening consequences if altered by unauthorized parties to be incorrect (e.g. an attacker changing the blood type of a patient in their records before a blood transfusion may lead to serious reactions due to blood incompatibility). This work suggests Distributed Ledger Technology (DLT) or Blockchain as a method of ensuring the integrity of healthcare data due to the immutable nature of this technology and compares the suitability of different variations of this technology for healthcare. Lastly, the current legal situation and regulations regarding the Blockchain technology and its applications is illustrated and areas where (further) regulations are needed as well as the deficiencies of the current regulations are identified.

2 Current State

Medical facilities today work with large amounts of sensitive data. There are several technologies such as cloud commonly used in order to be able to store and manage these large amounts of data and access them efficiently when needed. These data are also often protected through security measures (technical or otherwise) as they can easily be misused in the hands of malicious parties. But not only individual facilities, but also the government (and/or other policy makers) of each country is responsible for the protection of the medical data of its citizens through the use of data privacy laws and regulations and suitable punishment if these laws are broken or disregarded.

In this section, the current state of data protection in the healthcare sector will be discussed in detail. The first part of this section presents some of the current technologies widely used by medical facilities to collect, store and analyze and process medical data. The second section discusses the most common cybersecurity measures currently in use in the healthcare sector. Finally, the current data protection laws in the US, EU and EEA are discussed and examples of two different types of HIS, namely the chip-based system of Austria and the online healthcare system in Norway are compared in the third section.

2.1 Technologies and Trends in Healthcare

Developments in different technologies and different areas have caused an increase also in the use of many new technologies in healthcare during the recent years. For example, wearable IoT (Internet of Things) devices which improve the mobility of patients and facilitate their supervision are becoming more and more common in healthcare facilities today.

Internet of Things (IoT) is a sub-concept from the more broad and generic field of Ubiquitous Computing. The term "Ubiquitous Computing" was first introduced by Mark Weiser in his article "The Computer of the 21st Century" which was published in "Scientific American" in 1991 [7, p. 2]. Weiser worked at the Xerox Polo Alto Research Center which focused on researching and integrating the human factor into technology [7, p. 2]. This is important to note as at the time of Weiser's publication, operating and using a computer required a (relatively high) level of knowledge and specific skills which the majority of the common population did not have and so only a few limited number of people were able to use traditional computers then. The focus on this field at his workplace as well as the aforementioned situation at his time has possibly influenced and contributed to Weiser's research. Weiser in his article proposed the idea of having several computers connected to each other in a network with each computer focusing on one specific task. A computer in this context can be any object (e.g. a watch, fridge, plant pot, etc.) which has the ability to compute, meaning execute a given task through calculation or deriving an answer from acquired information. With this, the risk of a computer failing will be reduced with its reduced workload and the optimization and improvement of each task will be achieved more easily. The overall level of knowledge of each computer will still be the same as when one computer would do all the tasks as the computers will communicate and share information with each other through a network. The aim of Weiser was to integrate each of these computers (objects) into the user's everyday life in order to facilitate the use of computers for the majority of people without requiring a high level of background knowledge [7, p. 3]. This is also highlighted in his publication where he clearly distinguishes ubiquitous computing and virtual reality by saying that the former "brings the computer into the world" as opposed to the latter, which "brings the world into the computer" [7, p. 3].

Similar to ubiquitous computing, Internet of Things, from here on out referred to as IoT, is the concept of seamless integration of everyday physical or virtual objects or so called "things" in a network where all network members communicate and exchange data with each other [8, p. 26522]. As the name suggests, the focus with IoT lies on connecting each computer object with the internet and so using specifically internet as a network for the computers to communicate with each other (as opposed to using any possible network in ubiquitous computing). IoT devices are common in digital healthcare today and are used for supervising patients and collecting health-related data in real time [8, p. 26521]. There is extensive research on possible applications of IoT in healthcare such as monitoring patients with specific conditions such as Parkinson's disease or diabetes as well as supervising rehabilitation through monitoring a patient's progress [8, p. 26522-26523]. Additionally, the use of IoT devices provides the possibility to remotely monitor patients which could decrease the need for facility resources such as doctors, nurses or hospital beds and so reduce the pressure and (possibly high) workload of medical facilities [8, p. 26522]. Remote monitoring also allows those in need of medical care who live in remote rural areas to have better access to healthcare and for elderly or those with special conditions to live at home independently for longer [8, p. 26522].

The use of IoT in healthcare is still not very widespread and common, but it has been increasing quickly in recent years and has enormous benefits such as cost-reduction and increased efficiency [9, p. 678] so therefore it is appropriate that it is mentioned in this section. Additionally, the use of IoT allows one to improve the functionality of current medical devices and equipment, such as Glucometers (to measure the blood sugar level) or blood pressure or heart rate measurement devices. As discussed before, there is extensive (theoretical) research on possible applications using the currently available technologies such as sensors and wearable devices [9, p. 679]. There are however also some commercially available technologies which are currently already being used in the healthcare sector. The work of S.R.M Islam et al. [9, p. 690-693] provides an extensive overview of currently available and commercially in use IoT technologies some of which will be briefly summarized in the following. There are several wearable devices or sensors on the market, developed all over the world by companies such as Edisse, Garmin or Jawbone, which track their user's blood pressure, temperature and heart rate amongst others and offer an overall picture of the user's health status [9, p. 690]. These devices often have the ability to alert the user or medical staff in case of irregularities or medical emergencies [9, p. 690]. There are also products, for example a remote smartphone-linked door opener developed by LiftMaster, which facilitate and regulate home access and can be used by the elderly independently [9, p. 690]. An overview of a possible network architecture of these devices and how their communication and workload division may be structured is shown in figure 1. The figure shows a possible structure for collecting large amounts of health-related data such as vital signs. There are sensors which are directly worn by the user or patient (or placed in their body) which measure blood pressure or electric signals generated by the user's heart, brain or muscles through the use of ECG (electrocardiography), EEG (electroencephalography) or EMG (electromyography) sensors respectively. The electric signals of for example the heart, brain and muscles are recorded by the sensors which share these measurements with the computers (including not only traditional computers, but also for example smartphones or any other devices capable of computing as previously explained) or so called “resource providers” according to the figure. These devices provide the necessary resources (primarily for computing and storage) needed to process this data and can derive the user’s heart rate, brain activity rate and body temperature from the electric signals recorded by the sensors before and analyze and present the results. There may also sometimes be so called “brokers” which might facilitate and manage the workload and communication between the data providers and resource providers.

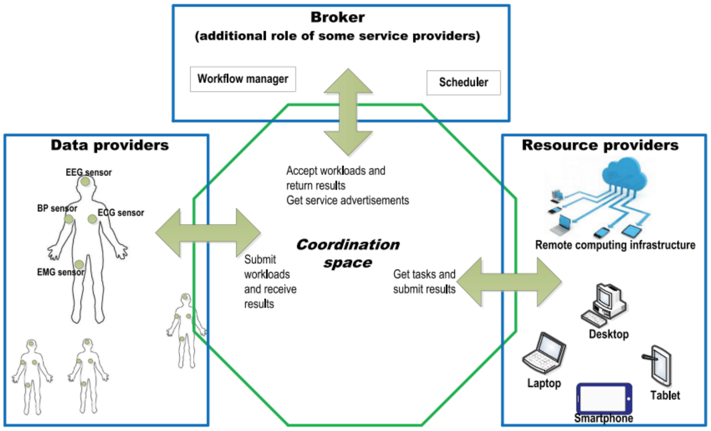


Fig. 1. A concept diagram of workings of IoT healthcare solutions [9, p. 680]

While IoT devices facilitate data collection and patient monitoring amongst other benefits, it is important to take great care when developing these devices and integrating them into an existing digital healthcare infrastructure. Simple issues with digital devices which wouldn’t be an issue in everyday life, such as running out of battery or being out of the network’s signal range, could prove to be fatal when concerning medical devices [8, p. 26522]. Additionally with regard to IoT specifically, it is important to ensure the security of the network being used for the computers communication and that no unexpected or unwanted devices can access and participate in this network as this might disrupt the communication between the necessary devices and even corrupt important patient data [8]. And there is also the security risk of storing large amounts of sensitive data in the same database but most of these problems are easy to solve and are already being improved with continuing research and further developments in this field [8, p. 26522]. It is most important to keep in mind not only the many benefits of IoT especially with regard to patient care in digital healthcare systems, but also possible negative aspects of unprepared applications and it should be considered that there are always areas to improve and possibility for continuous development of this concept and its applications.

As discussed above, IoT devices, amongst other functionalities and benefits, allow one to collect many relevant and important health-related data. Additionally, healthcare facilities all over the world have to store and handle millions of other health records and sensitive data daily. This is where two more common technologies, namely Cloud Computing and Big Data, come into the picture and can be utilized to respectively store and process data also in the healthcare sector.

Before the advancements in cloud computing, data would be stored and retrieved locally from a single machine (e.g. a computer) [10, p. 1]. This would lead to loss of important data in case of damage to the machine [10, p. 1]. To prevent this, the data had to be backed up and constantly updated in case of changes which in itself was a tedious process when done manually [10, p. 1]. Additionally, ensuring the security of the data on every single machine would be an issue and there were also limitations to the volume of the data a machine could store [10, p. 1]. Cloud computing, similar to other distributed systems, is defined as a network of multiple independent (remote) computers, data centers, servers or the likes which are integrated into the network and work together as a single facility to fulfill tasks such as storing and managing data [10, p. 2]. The benefit of cloud networks is that they provide users with the needed computing power and other services on-demand and are highly elastic, scalable, available and nowadays also relatively cheap [10, p. 2]. Scalibility means that the system can be easily expanded by statically adding resources (e.g. increasing the number of computers) if needed and without really improving or impacting the performance. Elasticity on the other hand is the rather dynamic ability to increase or decrease resources as needed to adapt to workload changes automatically and quickly and without any additional high costs to optimize and maximize the use of all resources. So scalability can be said to fulfill static needs such as an increasing amount of workload or data and elasticity is associated with dynamic changes in the workload such as increased amount of data around a certain time period which might later on decrease again. Availability in this context means that the system (and the data) are always accessible and (almost) never down when needed which is important especially in areas such as healthcare [10, p. 2].

These qualities make cloud computing flexible and reliable, and therefore also suitable for handling the increasing volume of health-related data as not all facilities have the ability to purchase and maintain resources for storage and processing of these large amounts of data. However, it is important to keep in mind that the availability of a cloud can never be fully guaranteed especially with the every day increasing amount of data [10, p. 11]. The article by Tahir Adnan et al. [10, p. 11-15] explains some strategies to reduce outages and possible data loss, the most common of which, namely Data Replication, Erasure Coding and Data Deduplication, are discussed in the following. As the name suggests, data replication is the practice to create multiple copies of the original data and to store these on one or multiple clouds [10, p. 11]. In this way, if one of the clouds has an outage or a specific data is inaccessible, the data will still be available on the other clouds (or a copy of it will be available on the same cloud) and can be accessed by the users [10, p. 11]. A strategy to replicate the data within the same cloud as well as on several other clouds is also mentioned in this article for additional security. The downside to this strategy is that the cost of data storage will increase with the increased amount of data which need to be stored [10, p. 12]. Therefore, it is recommended to additionally use cost reduction strategies such as load-balancing which means distributing the workload (e.g. user requests to access a file) between several available resources to optimize resource usage which leads to reduced costs [10, p. 12].

Erasure coding is the strategy to divide the data into a number of smaller blocks and to add some redundant data to each block in order to protect the original data in the in case of failures [10, p. 13]. This improves the availability of data and reduces the cost of storage compared to data replication [10, p. 13]. Through the redundant data, the original information can be reconstructed even if part of the data is lost which increases the system's tolerance for failures [10, p. 13]. Each of the blocks containing the original and redundant data is then stored on a different cloud storage so for example for a number n of blocks, there will be exactly n clouds, each storing one of the blocks respectively [10, p. 13]. There are some problems in some applications and implementations of this strategy especially regarding an increased CPU-workload which leads to increased time latency (slower access to data) [10, p. 14]. There are however certain methods such as pairwise balanced erasure coding design which optimize the distribution of workload and the retrieval of necessary data blocks for the reconstruction of lost data [10, p. 14].

In data deduplication, any redundant data or copies are eliminated and only one instance of each data remains on the cloud which has that data stored [10, p. 14]. Other clouds or backups will only have a pointer reference to the original instance [10, p. 14]. A unique hash value for the identification of each data which serves as the pointer reference or so called fingerprint of that data will be created using a chosen hash algorithm[10, p. 14]. Then, duplicate data with the same hash value or fingerprint will be removed from the cloud(s) and the rest, namely the unique data, will be stored with their corresponding hash value [10, p. 14]. Storing all of hash values for later comparison can become problematic for larger numbers of data [10, p. 15]. Some solutions here would be to group hash values of the same application together (and then later on divide it into groups depending on the application) or to additionally implement a fingerprint index management which uses the hash values as an index and so eliminates duplicates with minimum use of resources such as read or write bandwidth [10, p. 15].

According to Tahir Adnan et al.[10, p. 16], data replication seems to be the most common practice today and is also the one mostly suggested by researchers. However, with the increasing volume of data, there is increased concern about their availability and security as well as an increased need for more efficient strategies compared to simply increasing the number of data replicas which can lead to increased cost and an increase in the required storage space in the long term [10, p. 20]. There are several storage mechanisms proposed by different researches which use one or a combination of the strategies discussed above to improve data availability. A complete list of these strategies can be found in the aforementioned article [10, p. 16-19].

Collecting very large amounts of data has become incredibly easy today through the use of consumer technologies such as the wearable devices for patients mentioned before. With increased technological advancements in fields such as IoT, these technologies have continued to be better integrated and used more often in many areas including healthcare in the recent years. This has lead to a significant increase in the volume of data collected from for example patients and therefore a significant increase in the overall amount of medical data which will only continue to grow in the future. According to the article by Abhinav Rai [11] from 2020, "businesses around the world generate nearly 2.5 quintillion bytes of data daily almost 90% [of which] has been produced in the last two years alone". This signifies that the volume of collected data will only continue to grow exponentially and with increasing speed in the future. These large, complex and exponentially growing amounts of unlinked data which are very difficult or impossible to process with current data processing tools or methods are also known as "Big Data" today [11]. Big data is often defined as having high variety (several different forms structured or otherwise such as pictures, audios, documents, etc.), velocity (the speed at which data is collected or created in real-time) and volume (huge volume generated by several sources) [11]. Big data can be divided into structured, unstructured and semi-structured data [11]. Structured data is the easiest of the three to work with as it is stored and can be retrieved and processed in a fixed format which is uniform for all stored data [11]. Some of the current search algorithms and simple data processing tools even work on these highly organized (e.g. in a table) data sets [11]. Unstructured data on the other hand lack any structure or organization or a specific format which makes it more difficult and time consuming to process these [11]. Semi-structured refers to data that contains both structured and unstructured formats, meaning that although the data is not explicitly organized in e.g. tables, it still contains information (such as XML tags) that mark individual elements within the data which facilitate the processing of these compared to completely unstructured data types [11].

As previously mentioned, it is possible to store and preserve big data sets as well as ensure that they are accessible almost any time they are needed using cloud computing and certain availability strategies. The processing of large volumes of big data (especially unstructured data) such as ones collected by the wearable devices mentioned before, however, cannot be achieved very efficiently with cloud computing. For this purpose, concepts and technologies such as Machine Learning and Artificial Intelligence have been suggested to be suitable by researchers and implemented to a certain degree in healthcare today. The two expressions are often used interchangeably, it can however be argued that there are minor differences. Artificial Intelligence is the overall ability of a computer or a machine to use reasoning to solve problems and fulfill given tasks on its own. The most important characteristic of artificial intelligence is that the computer should have the ability learn independently without the need for (constant) human intervention and support. Machine learning can be defined as a collection of learning methods and algorithms through which the computer or machine can be trained and "develop its own intelligence" to use for learning and to acquire new knowledge for future tasks and problem solving. There are several mathematical algorithms such as clustering, where similar data or values are grouped together in so-called clusters, which can be used for the purpose of training a computer. So machine learning can be said to be a more specific area and a practical application of the general concept of artificial intelligence.

Some abilities of machine learning, from here on out referred to as ML, that have direct applications in healthcare are recognizing patterns in data, predicting future outcomes based on previous patterns as well as image or object recognition and classification [12, p. 525]. Advancements in artificial intelligence, from here on out referred to as AI, and ML have made it possible to deduce associations, correlations and causation in large complex data sets such as big data which are unstructured and non-normalized [12, p. 525]. An overview of the possible applications of AI in healthcare can be seen in figure 2. AI diagnostic systems today rely on ML to identify patterns which would be unrecognizable to humans and could therefore not be detected before [13, p. 224]. An example of one such diagnostic system is "the Deep Patient initiative in which a research group at Mount Sinai Hospital in New York trained a program using the electronic health records of 700,000 patients and then used the program to predict disease in another sample of 76,214 patients" [13, p. 224]. The results of the diagnostic system were said to significantly outperform the results of raw health record data analysis or alternative future learning strategies[13, p. 224]. While this research and similar practical applications have yielded significant and mostly positive results until today, it is important to keep in mind that such deductive AI systems can also be biased and deliver inaccurate results. For example, it has been observed that most algorithms in dermatology have been trained with Caucasian or Asian patients and may produce inaccurate results if used on patients with any other ethnicity [13, p. 226]. While it is difficult to completely prevent this bias, it can be reduced significantly by training the program with a large amount of data which is representative of the future patients it might be used on and also additionally ensuring that the algorithm used to analyze those data is objective and free of any biases its developers might have [13, p. 226]. It is important to note that even small biases may be reflected several-fold in the results that the AI yields in the long term as it learns from its own previous results over and over again.

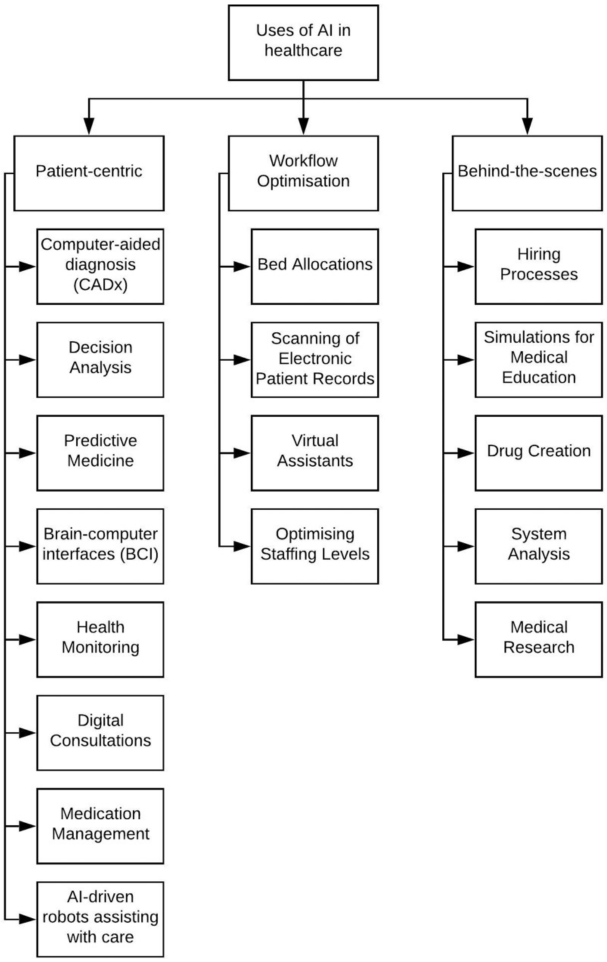


Fig. 2. Overview of possible AI applications in the healthcare industry[13, p. 225]

Similar to pattern recognition, the program also has to be trained with a set of data for image recognition. An example of this is the use of a large amount of labelled cat and dog pictures to train a computer to develop its own methods for recognizing and distinguishing between cats and dogs in pictures in the future. AI is already proving capable of outperforming human clinicians in the diagnosis of specific medical conditions especially through image analysis in the fields of dermatology, cardiology and radiology. H.A. Haensle et al. trained Google's

Inception v4 CNN architecture with pictures of dermoscopic images and corresponding diagnoses and then used it to detect melanoma detection (a type of possibly fatal skin cancer) which has been increasingly fatal in the recent years [14, p. 1836]. The results yielded by the trained CNN were compared to 58 international dermatologists (30 of which were experts) and most of them were outperformed by the CNN [14, p. 1838-1839]. Similarly, Andrew J. Steele et al. used approximately 82,000 electronic health records in total to train and test a program to predict the mortality risk of individual patients with coronary artery disease [15, p. 8] relatively successfully and concluded that it was more efficient (time-wise) in suitable variable selection for chosen statistic methods than a human would be and that unstructured and missing information in health records could also be handled relatively easily and quickly with minimal human intervention [15, p. 14]. This means that there is great potential to reduce the workload of medical staff (which are already exhausted due to staff shortage in certain fields) as such systems can work for longer hours and are partially also more efficient in certain areas compared to human workers. Finally, Pranav Rajpurkar et al. developed an algorithm for detecting Pneumonia from chest X-Rays by using over 100,000 frontal view X-Ray images with 14 diseases which was observed to outperform 4 practicing radiologists on the average precision and performance on 420 images [16, p. 1-2]. Overall, all three experiments concluded that the medical staff and experts, regardless of their expertise and experience level, could benefit from the assistance of such deductive AI systems. Such systems can help transform the healthcare industry from often being reactive only once diseases have advanced to a serious stage to a more preventive nature which allows earlier discovery of symptoms and diagnoses. They can also predict for example the possible future reactions of a patient to a specific treatment based on previous patterns and thus can often significantly reduce fatality and burden of diseases and allow medical staff to focus on individual patients rather than generic statistical indicators.

In addition to the deductive systems used for analyzing data and images and diagnosing, there are also generative AI systems which can create synthetic mock patient data once they have been trained with an existing set of digital health records [13, p. 224]. This can be used as training material for medical staff to further improve their knowledge and experience as well as their ability to recognize and diagnose diseases more early on which may indirectly lead to better patient outcomes [13, p. 224]. However, it is important to keep the weaknesses of current AI systems, such as possible biases mentioned above, as well possible consequences of extensive use in mind. While AI can assist healthcare professionals tremendously in diagnosing and treating patients, there is also the possibility of the staff developing automation bias, meaning that they become overdependent on these systems and trust their results over their own professional medical opinion [13, p. 228]. Because of this, it is crucial that all staff especially doctors working with deductive AI are able to understand how it works and comprehend its decision-making process [13, p. 228]. If this is achieved, then the staff are always able to critically assess the results the AI program or system has produced and intervene or overrule its results if they for example think that its diagnoses aren't accurate. Currently, documenting the details of the decision-making process of AI systems can be difficult due to these systems teaching themselves patterns [13, p. 228]. This lack of transparency of AI methods is referred to as the "black box effect" which limits clinical use unless more projects which clearly illustrate the decision-making process of the ML system used, such as the system developed by researchers at DeepMind and Moorfields Eye Hospital to interpret optical scans and treatments, are created [13, p. 228].

AI does not only have the potential to support medical staff, it can also enhance existing technologies such as IoT and medical devices such as digital monitoring equipment and allow them to work more dynamically and respond to individual patient needs [13, p. 223]. Therefore, it is important that AI applications are thoroughly integrated into the existing technologies and IT-infrastructure also in the healthcare sector [13, p. 223]. This might require considerable changes to the existing infrastructure which can be costly and can demotivate organizations due to financial or functional reasons such as the high amount of time it might take to not only implement, but also fully integrate the new technology and educate the staff about it. Thus, there will be risks such as automated and systematic bias, over-dependence and possibly (medical) data privacy concerns unless the introduction of AI is done gradually and after careful consideration and examination [13, p. 223]. With the increased potential of AI and the increased possibility of its integration in the near future, it is also becoming a topic of interest to regulators [13, p. 224]. It is vital for governments and authorities of each country to regulate and control the applications of AI, specifically in their healthcare systems, at least to a certain extent especially given that healthcare organizations around the world are most likely not perfectly prepared for this transition currently.

2.2 Cybersecurity Measures

The use of new technologies and advancements and improvements in existing fields can be a big help for a more efficient handling of medical data if done correctly. Medical facilities today use several technologies, amongst others the ones mentioned in the previous section 2.1, to facilitate and improve their works and services. However, as briefly mentioned before, there are always security concerns as it is often the case with technologies in practice. These concerns increase for new technologies as not all weaknesses of these are known due to unfamiliarity with them. Insufficient or missing implementation and integration of technologies also leads to more security breaches which can be misused if not discovered and resolved properly. A breach means that information and data, in this case patient records, are lost, stolen, displaced, hacked or communicated to unauthorized parties [17, p. 2].

In order to be able to defend a system from threats, it is of utmost importance to first identify and understand the threats to that system. Cybersecurity attacks are classified based on different criteria such as their purpose or severity [17, p. 2]. An article by S.S. Bhuyan et al. [17, p. 2-3] identifies Denial of Service (DoS), Privilege Escalation, Man in the Middle (MITM), cryptographic attacks, SQL Injection as well as the use of malicious software and phishing as major types of cybersecurity threats to healthcare organizations. Denial of Service attacks aim to flood a network with several requests in order to disrupt its functions and deny the users any service and access to the network's resources due to the network slowing or even completely shutting down because of the large amount of traffic [17, p. 2]. Privilege escalation attacks exploit security vulnerabilities to gain privileges other than what was originally intended for the user [18, p. 110]. Such attacks can be horizontal, where a user tries to access the account of another user with the same privileges, or vertical, where a user tries to exploit a flaw in the system to gain higher privileges and more access rights than was intended for them [18, p. 110]. Horizontal privilege escalation can for example lead to a (hacker) patient accessing not only their own medical records and information but also those of another patient which breaches the other one's privacy and could possibly endanger them [18, p. 110]. A vertical privilege escalation would for example include a user gaining admin rights to a network through exploiting a vulnerability of that network [18, p. 110].

In Man in the Middle attacks, the attackers act as an intermediary in the conversation of two parties [17, p. 2]. If successful, the intruder can not only intercept the messages and information exchanged between the parties, but also alter the data from one party before relaying it to the other party without either one knowing the data has been compromised [17, p. 2]. In case of medical records, this information leakage could endanger patients and their privacy and possibly even lead to blackmail. Cryptographic attacks aim to decrypt encrypted information without authorization [17, p. 2]. Sensitive information are sometimes encrypted, meaning altered to a non-comprehensible format for anyone other than the sender and the receiver, in order to prevent unwanted readings. A more detailed explanation of different encryption possibilities will follow later in this section. The programming language ”Structured Query Language”, often abbreviated as SQL, is used in many websites to access and manage their database [17, p. 2]. An attacker could exploit any vulnerabilities in the SQL source code of a website, for example by using harmful SQL statements in the search bar of a website, to get access to and even alter or delete information [17, p. 2]. This is one of the simpler types of attacks but it can also easily be prevented through following coding standards and testing for possible obvious vulnerabilities before deploying an application or website.

The aforementioned attacks were mostly of technical nature, however the human aspect is also very important and should be considered when assessing threats. The human aspect includes all possible human-errors, for example the employees of a facility giving away sensitive information or unknowingly granting access to hackers due to insufficient or missing education. According to the summary of the cyberattacks and events of interest made by HHS (US Department of Health and Human Services), Ransomware and Email Phishing were among the most common cyberattacks on the health industry in 2021 [19, p. 11-57]. Both phishing and the spread of ransomware are results of direct or indirect human-error and the exploitation of the human factor (people). Malicious software or malware refers to any type of program designed to compromise or harm a computer system by altering, damaging, spying or deleting user information without the permission of the user [17, p. 3]. Malware can spread physically, for example through a corrupted USB-Drive which a hospital employee might receive from a stranger and stick in their work computer, or through downloads, for example by clicking the link on an email and downloading a corrupt file as a result of that [17, p. 3]. Ransomware, a type of malware which encrypts user's information and demands ransom in exchange for decrypting them and allowing the user access, specifically is often used to attack facilities today [17, p. 3]. Phishing is defined as the use of social engineering to manipulate and trick users into divulging sensitive information or performing harmful activities to their computers often unknowingly and it is also one of the most common ways to deliver malware [17, p. 3]. The case mentioned previously with the employees receiving an email containing a link to download a malware is an example of phishing.

Attackers might also combine several types of attacks to increase their chance of success. A well-known example of this is the case of the attack on Boston Children's hospital in 2014 by the hacktivist group "Anonymous" [20, p. 1]. The attackers first attempted to shut down the hospital network at the beginning of April with a relatively low rate of malicious traffic, then later on increased the rate of requests significantly over the course of the month and combined the DoS attack with other types such as SQL injection and phishing emails [20, p. 2]. After noticing the initial threat, the incident response team of the hospital was notified and had to assess the situation from a business, technical and clinical perspective and defend the hospital's resources and data [20, p. 3]. One of the organizational weaknesses in this case is the lack of previous threat assessment and that there was no analysis of what resources or services could be suspended in case of such attacks [20, p. 3].

Being aware of the most common threats and types of attacks, it is now important to identify different stakeholders and involved parties in cybersecurity who play a major role in either ensuring the security of data or jeopardizing it. Identifying all involved parties and understanding their roles as well as their limits will help with better planning to prevent security breaches [17, p. 4]. According to the article by S.S. Bhuyan et al. [17, p. 4], there are four major parties involved here: Attackers, end-users, defenders and developers. Cyber-attackers are the main threat to cybersecurity and the reason why it exists to protect valuable data and information from their attacks [17, p. 4]. The type of attacker is determined based on the intentions and authorization status of the attacker [17, p. 4]. The attacker, or hacker, is an individual that seeks to gain remote access to a system and data. Some hackers are hired and/or authorized to attempt attacks on a system and do not have any malicious intent [17, p. 4]. This is known as ethical hacking. Some researchers classify attackers into hacktivists, terrorists, spies and criminals [17, p. 4]. Hacktivists such as the previously mentioned "Anonymous" group are motivated by non-monetary ideals and aim to promote their political agendas through their cyber attacks [17, p. 4]. Cybercriminals on the other hand use a computer to commit crimes such as extortion or theft for monetary benefits [17, p. 4]. Cyberterrorists are defined as hackers who purposely aim to disrupt computer networks while those who engage in espionage concerning classified or proprietary data are called cyberspies [17, p. 4]. End-users can either be malicious or non-malicious, and even non-malicious users such as employees with insufficient or lacking security training can aid in attacks unknowingly [17, p. 4]. "A study of over 900 breaches in 2010 revealed that insiders who are either current or former employees were responsible for orchestrating 48% of all data breaches in the study, and only 10% of the incidents were unintentional" [17, p. 4]. So proper employee training and education as well as control and managing of access rights of the end-users is crucial. This will be explained in more detail in the coming parts of this section.

Cyber-defender includes all individuals, e.g. IT-professionals or government agencies, who ensure cybersecurity [17, p. 4]. The primary role of these defenders is to plan and execute security measures to ensure their organizations are protected from cyber threats and in case of some government agencies to additionally apprehend and charge cybercriminals [17, p. 4]. The next section 2.3 discusses the role of government agencies and some of their policies regarding cybersecurity and the protection of data in more detail. Developers are responsible for programming, and it is often their mistakes that are exploited by attackers [17, p. 4]. According to HHS, software vulnerability, along with the aforementioned phishing and ransomeware, is among the most common attacks in healthcare in 2021 [19, p. 11-57]. In order to reduce this risk, it is important for the developers to also be educated and aware of possible security risks and program with security in mind and while following standards and communicating and working closely with defenders.

According to K. Abouelmehdi et al. [21, p. 75-76] the most widely used strategies for protecting medical records and data in healthcare currently are authentication, access control, encryption, data masking or a combination of these. Similar to other branches, authentication is used in healthcare to confirm and verify the identity of users of devices with access to an organization's network (e.g. hospital computers) [21, p. 75]. The increase in the use of portable devices, such as the IoT devices mentioned in the previous chapter 2.1, in different areas such as patient care and other applications has caused an increased need for authentication. The use of mobile devices such as tablets or work smartphones allows individuals to perform their daily tasks from anywhere at any time [22, p. 1] and increases the availability of medical staff. But this also means an increase in the number of devices and nodes that have to be secured and kept safe which leads to more exposure and a higher risk of attacks to the security of an organization's network and confidential data. There have been several traditional authentication methods proposed by researchers, which can generally be sorted into three different categories [22, p. 1]:

1. Knowledge-Factor: Information that the individual authorized to use a device/service should have. This requires the user to answer some questions which ideally only the authorized individual should know the correct answers to [22, p. 1]. Some well-known examples of this are passwords or PINs [22, p. 1].

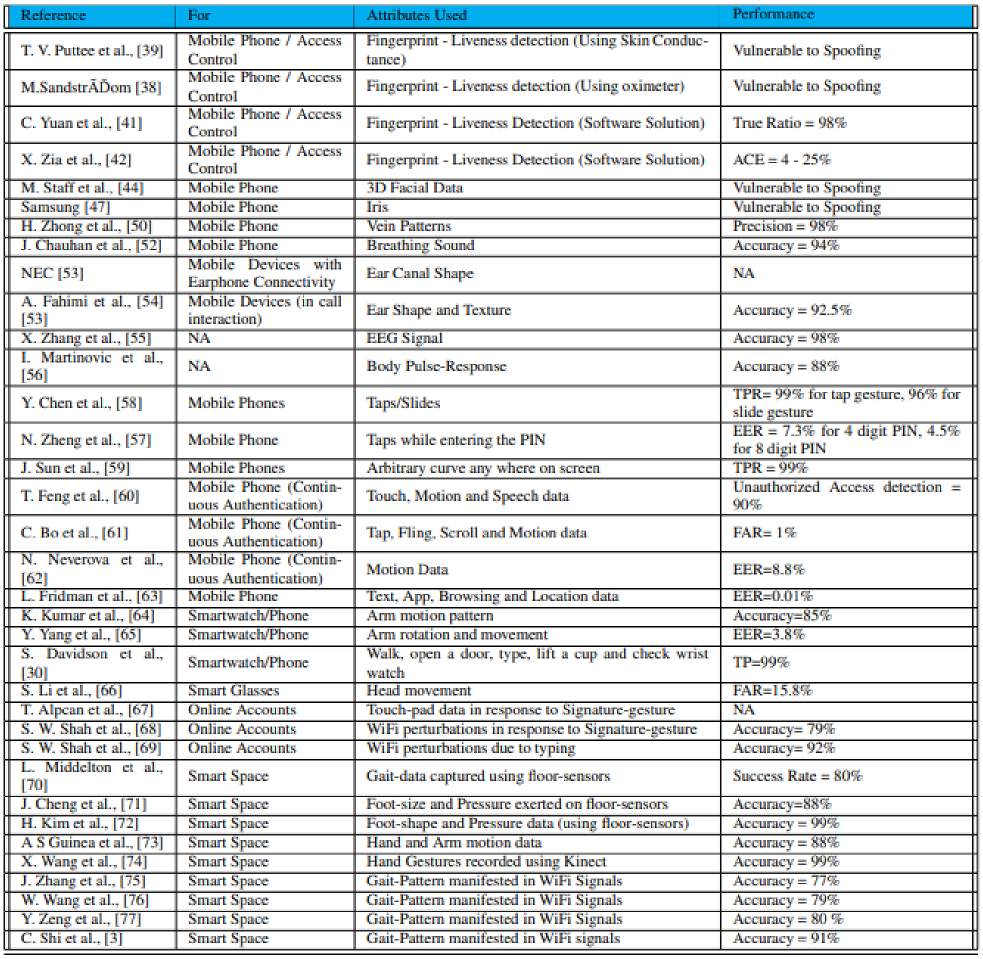
2. Inherence-Factor: Something that the individual is. This category often relies on physical or behavioral biometrics. Physical biometrics such as fingerprints are based on one's physical trait [22, p. 2]. Behavioral biometrics, such as the way a person walks, establish an individual's identity by identifying unique patterns in their behavior [22, p. 2]. Inherence-based authentication generally requires an enrollment procedure, where the user provides several samples of their physical or behavioral trait [22, p. 2]. The authentication algorithm then extracts and stores some features of the samples which allows it to distinguish and authenticate the same traits in the future [22, p. 2].

3. Possession-Factor: An object that the individual owns. This method relies on the authorized user possessing some form of a hardware such as the widely used RFID swipe cards to restrict access to certain restricted places [22, p. 3].

However, all of the mentioned categories have been proven to have certain vulnerabilities. In the first category, users tend to choose very simple passwords that are easy to guess and there are also several lists of the most used passwords published online which facilitate brute force attacks for hackers [22, p. 1]. Although there has been a 10% decrease in the success rate of brute force attacks recently due to users being more security-aware and choosing more complex passwords, there are still other methods such as social engineering, malware, leaked password databases or even simple oil smudges on the screen for compromising even more complex passwords [22, p. 2].Additionally, it is also noteworthy that a single user today often has several online accounts (social media, payment platforms, online shops, etc.) and it is difficult to come up with a different strong password for each of those so users might use the same password for several accounts, meaning that all of those will be compromised if the password of one of those is hacked [22, p. 2]. Password managers help with this problem to some extent but even they themselves can be compromised [22, p. 2]. In the second category, face-recognition and fingerprint IDs have become common in the past few years but these can be easily compromised by getting the fingerprint of the individual from any surface they touch or just using a picture of the individual (easily found for example on social media) to trick the fingerprint or face recognition algorithm [22, p. 2]. Another big problem is that these authentication methods have no recourse if compromised; for example once a hacker gets the fingerprint of an individual from a surface, they can use that fingerprint in the future as well unlike passwords which can be changed several times with no problem if compromised [22, p. 2]. Researchers have been trying to improve the algorithms used here to for example detect whether the finger used for fingerprint recognition is alive or a fake but even these methods can still be fooled in several ways and so the research here is still ongoing and not implemented in the mainstream commercial products [22, p. 8]. A summary of some of these methods in development can be found in figure 3.

Lastly, possession-based authentication methods also have downsides and may also be susceptible to hacking. The reliance of these methods on physical hardware means that the users must be carrying the corresponding hardware with them all the time which can be bothersome for the users and they may also forget the hardware or it can be easily stolen or copies by non-authorized parties and hackers [22, p. 3]. Additionally, it is also costly for the service providers to provide all users with the requires hardware [22, p. 3]. Nowadays, service providers have begun to use smartphones instead of separate pieces of hardware, e.g. One Time Codes generated on online banking apps, but even these require extra user interaction to for example copy the code from the app into a shop's website to confirm authentication when purchasing something which limits this approach [22, p. 3]. As it can be seen, each category has its own limits and can be deceived sometimes despite being highly accurate the rest of the time. A solution which is also on the rise nowadays is to use (two- or) multi-factor-authentication which combines methods from two or more of the aforementioned categories [22, p. 3]. This combines the strengths of each category and provides backup security in case one of the methods in use is compromised or fails. It also makes it more difficult for non-authorized users or hackers to hack accounts.

Once a user is authenticated, there may be access control policies which regulate the information a user has access to and what activities a user can perform. The most common policies in healthcare are Role-Based Access Control, where a user's access rights are defined by an assigned role (e.g. nurse), and Attribute-Based Access Control, where not only the role, but also other attributes such as the action to be performed (e.g. reading a file) are considered [21, p. 76]. Data encryption means that the data is converted to an unreadable set of characters, which can only be decoded and read using a special key held by authorized parties. There are several encryption algorithms which are used by different organizations. Organizations should choose a suitable algorithm which is effective but also easy to use for both patients and the medical staff [21, p. 75]. Similarly, data masking replaces sensitive data such as patient name or birthdate with an unreadable expression before further deploying the data [21, p. 75]. The difference to encryption is that the original patient information (e.g. name or



Reference	For	Attributes Used	Performance
T. V. Puttee et al., [39]	Mobile Phone / Access Control	Fingerprint - Liveness detection (Using Skin Conductance)	Vulnerable to Spoofing
M.SandstrÃĎom [38]	Mobile Phone / Access Control	Fingerprint - Liveness detection (Using oximeter)	Vulnerable to Spoofing
C. Yuan et al., [41]	Mobile Phone / Access Control	Fingerprint - Liveness Detection (Software Solution)	True Ratio = 98%
X. Zia et al., [42]	Mobile Phone / Access Control	Fingerprint - Liveness Detection (Software Solution)	ACE = 4 - 25%
M. Staff et al., [44]	Mobile Phone	3D Facial Data	Vulnerable to Spoofing
Samsung [47]	Mobile Phone	Iris	Vulnerable to Spoofing
H. Zhong et al., [50]	Mobile Phone	Vein Patterns	Precision = 98%
J. Chauhan et al., [52]	Mobile Phone	Breathing Sound	Accuracy = 94%
NEC [53]	Mobile Devices with Earphone Connectivity	Ear Canal Shape	NA
A. Fahimi et al., [54] [53]	Mobile Devices (in call interaction)	Ear Shape and Texture	Accuracy = 92.5%
X. Zhang et al., [55]	NA	EEG Signal	Accuracy = 98%
I. Martinovic et al., [56]	NA	Body Pulse-Response	Accuracy = 88%
Y. Chen et al., [58]	Mobile Phones	Taps/Slides	TPR= 99% for tap gesture, 96% for slide gesture
N. Zheng et al., [57]	Mobile Phone	Taps while entering the PIN	EER = 7.3% for 4 digit PIN, 4.5% for 8 digit PIN
J. Sun et al., [59]	Mobile Phones	Arbitrary curve any where on screen	TPR = 99%
T. Feng et al., [60]	Mobile Phone (Continuous Authentication)	Touch, Motion and Speech data	Unauthorized Access detection = 90%
C. Bo et al., [61]	Mobile Phone (Continuous Authentication)	Tap, Fling, Scroll and Motion data	FAR= 1%
N. Neverova et al., [62]	Mobile Phone (Continuous Authentication)	Motion Data	EER=8.8%
L. Fridman et al., [63]	Mobile Phone	Text, App, Browsing and Location data	EER=0.01%
K. Kumar et al., [64]	Smartwatch/Phone	Arm motion pattern	Accuracy=85%
Y. Yang et al., [65]	Smartwatch/Phone	Arm rotation and movement	EER=3.8%
S. Davidson et al., [30]	Smartwatch/Phone	Walk, open a door, type, lift a cup and check wrist watch	TP=99%
S. Li et al., [66]	Smart Glasses	Head movement	FAR=15.8%
T. Alpcan et al., [67]	Online Accounts	Touch-pad data in response to Signature-gesture	NA
S. W. Shah et al., [68]	Online Accounts	WiFi perturbations in response to Signature-gesture	Accuracy= 79%
S. W. Shah et al., [69]	Online Accounts	WiFi perturbations due to typing	Accuracy= 92%
L. Middelton et al., [70]	Smart Space	Gait-data captured using floor-sensors	Success Rate = 80%
J. Cheng et al., [71]	Smart Space	Foot-size and Pressure exerted on floor-sensors	Accuracy=88%
H. Kim et al., [72]	Smart Space	Foot-shape and Pressure data (using floor-sensors)	Accuracy = 99%
A S Guinea et al., [73]	Smart Space	Hand and Arm motion data	Accuracy = 88%
X. Wang et al., [74]	Smart Space	Hand Gestures recorded using Kinect	Accuracy = 99%
J. Zhang et al., [75]	Smart Space	Gait-Pattern manifested in WiFi Signals	Accuracy = 77%
W. Wang et al., [76]	Smart Space	Gait-Pattern manifested in WiFi Signals	Accuracy = 79%
Y. Zeng et al., [77]	Smart Space	Gait-Pattern manifested in WiFi Signals	Accuracy = 80 %
C. Shi et al., [3]	Smart Space	Gait-Pattern manifested in WiFi signals	Accuracy = 91%

Fig. 3. Summary of experimental inherence-based authentication methods currently in development to improve accuracy and security [22, p. 8]

birthdate) cannot be derived from the masked value [21, p. 75]. This anonymization technique is often used as it reduces the cost of deploying large amounts of data [21, p. 76].

However as mentioned before, the human aspect of cybersecurity is also very important as human errors are one of the leading causes of security breaches. "Ponemon Institute reports that at least 78% of data security breach is the result of human negligence or maliciousness" [23, p. 1] and "IBM Global Technology Services 2014 declares that 95% of information security incidents involve human errors" [23, p. 2]. Sources cited in [23, p. 2] argue that healthcare systems involve technology as well as people so security threats cannot be prevented by a merely technical approach, but instead by a multi-disciplinary approach consisting of all aspects of the society including people. Thus, correct and sufficient education of all employees of a medical facility on security threats and the consequences of their actions among others is crucial in ensuring that common social engineering methods or phishing and spreading of malware do not work or at least have a reduced rate of success and so ensuring the security of the sensitive data and medical records often stored in these facilities [23, p. 1]. Information security training programs today are common in different organizations in the healthcare branch to educate all employees on security-aware behaviors such as updating Software and systems regularly to avoid vulnerability and recognizing phishing emails. However, some security awareness trainings fail to have a long-term effect on the employees, as the programs are not engaging and do not encourage creative activities and critical thinking [23, p. 2]. Training should not only be engaging, but also be done regularly and be designed with the level of technical knowledge of the participants in mind [23, p. 2]. The most important content to include in the training is the organization's internal security policy in addition to the major threats to the organization information assets as well as basic safeguards (e.g. choosing a strong password) and incident management [23, p. 2-3]. The training content should help employees to not only recognize possible security issues but also to respond accordingly [23, p. 3] to prevent the most amount of damage possible. Generally, making the trainings more interactive (and easy to follow), engaging and regular and measuring the progress of the participants and improving or changing the training based on that will improve the users' education and security-awareness [23, p. 2] and thus prevent or reduce considerable amounts of security breaches in the long run.

All in all, it is important to note that it is impossible to guarantee the cybersecurity of an organization's system and to avoid all security breaches forever. Security-awareness programs and physical or technical barriers are mainly there to reduce the success rate of cyberattacks and to ensure that the damage done by the breaches (e.g. number of leaked patient profiles) is minimal if successful. Policymakers and the government of each country have and will continue to enforce policies and laws to help healthcare organizations to protect their citizen's sensitive medical data which will be discussed in more detail in the next chapter 2.3.

However, these policies are often not sufficient and take a long time to enforce or change so it is recommended that healthcare organizations also implement some security measures of their own. Implementing and integrating any security protocol, technical or otherwise, requires both financial and other resource (e.g. time) and involves trade-offs as no company can have infinite resources. Thus, a careful cost-benefit analysis should be done to find the most suitable approach as this differs for each organization.

2.3 Current data privacy regulations

As briefly mentioned previously, not only healthcare organizations, but also the governments and policymakers of each country are responsible for regulating and protecting healthcare cybersecurity and the sensitive data of their citizens. Policymakers face an ever-changing and evolving concept when dealing with cybersecurity. Changes in digitalization trends (also in healthcare) such as an increase in the number of mobile devices require an accordingly flexible and quick response from the policymakers. This is, however, not always possible as establishing and enforcing new regulations takes time and can be difficult to change so policymakers might constantly have to try to catch up to the trends and regulate them only for the next trend to emerge quicker than they are able to establish new regulations. On the other hand, policymakers can also initiate trends and encourage and support suitable innovations by issuing new regulations or improving existing ones to encourage the integration of emerging technologies in (healthcare) organizations and to ensure their security.

In state-funded healthcare systems, the government's role is balancing the desire for the highest possible care standards with strict financial constraints [13, p. 227]. Relevant government policy and regulation is often influenced heavily by public opinion with input from stakeholders within the health service [13, p. 227]. Thus, even when IT systems indirectly improve patient outcomes, investment may be viewed less favourably than for items which directly improve patient care with quickly visible effects [13, p. 227]. For example, cutting investment into staffing and equipment in order to invest in a new technology (unknown to public and stakeholders or disliked by them) could be met with an unfavourable public response [13, p. 227]. In this way, it is not enough for technologies to simply be proven that they are capable of improving current healthcare standards, they must gain the advocacy of key stakeholders and be acceptable to the public [13, p. 227].

Different countries each have their own (increasingly digital) healthcare system and so accordingly a different set of laws and regulations regarding data privacy as well as different enforcers. In the US, for example, there are several different departments within the government, such as the Department of Homeland Security and the Department of Justice, which are responsible for apprehending and charging cybercriminals who do not follow the cybersecurity regulations [17, p. 4]. Some agencies like the National Institute of Standards and Technology contribute to the development of frameworks for ensuring cybersecurity and even the Congress in the US has taken an active role in developing laws to mitigate cybercrime, having enacted at least six cybersecurity bills between the 113th and 114th Congress [17, p. 4]. As healthcare organizations transition to electronic-based systems, many are left vulnerable to cybercrime. "Cybercrime emerged in the late 1970s as the Information Technology industry took shape. What began as spam eventually transitioned into viruses and malware" [24, p. 1] and today the attacks and the technology used by hackers are only becoming more sophisticated and coordinated. In order to circumvent the breach of healthcare data, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) implemented physical, technical and administrative safeguards to ensure sensitive information is protected from cybercriminals [24, p. 2]. Although HIPAA was passed by the Congress in 1966, its sub-rulings regarding security were only enforced, meaning that compliance with them became a legal obligation, starting the 20th of April 2005 for most covered entities and only in starting the 23th of September 2013 for business associates [25, p. 1]. HIPAA defines three pillars to securing the protected health information which is included in a patient's electronic health record [25, p. 2]:

1. Physical safeguards: techniques that prevent or limit physical access to a resource to only allow access by authorized parties. For example the RFID card of a front-desk clerk will not open the emergency room as they typically would not need access to there or patients are only limited to their own ward and the entrance to the other wards are blocked by card scanners in a hospital [25, p. 2].

2. Technical safeguards: collection of methods that prevent or limit access to electronic resources. Examples of these are the aforementioned access-control policies or encryption and automatic log off and emergency access protocols [25, p. 2]. The idea of using a unique patient identifier number that can map to a number of data sets collected by the government (thus keeping the identity of patients secret) was also first mentioned in the HIPAA Act as a technical safeguard.

3. Administrative safeguards: techniques that do not entirely belong to the physical or technical safeguards. These safeguards are typically in the form of policies or practices to for example regularly check for vulnerabilities and to continually improve the security of the organization's systems [25, p. 2].

These three pillars of security, also known as the three themes of security safeguards, and their applications and uses according to some works of literature are illustrated in figure 4. On the 17th of February 2009, the Health Information Technology for Economic and Clinical Health (HITECH) Act was signed into law by President Barack Obama which aims to encourage the adoption of health information technology and specifically the use of electronic health records by all organizations in the US [26]. The act was enforced starting the

30th of November 2009. Prior to the passing of the HITECH Act, only 10% of hospitals had adopted EHRs which made care coordination and communication and health information exchange between different hospitals difficult [26]. The act provided incentives to help with the high costs of transitioning from paper records to EHRs and thus increased the rate of adoption of of EHRs from 3.2% in 2008 to 14.2% in 2015 [26]. It made the implementation and use of EHRs mandatory for the recipients of federal funds such as Centers for Medicare and Medicaid Services (CMS) in order to receive full reimbursements, and those who had not implemented EHRs by 2015 would be penalized [25, p. 2]. By 2017, 86% of office-based physicians had adopted an EHR and 96% of non-federal acute care hospitals have implemented certified health IT in the US [26].

The HITECH act also aims to improve and further specify the language of the HIPAA act and introduced tougher penalties such as high fines for failure to comply with HIPAA to add an incentive for health organizations and their business associates to follow the HIPAA regulations [26]. HITECH additionally underlines the importance of reporting data breaches and requires a certain protocol to be followed to do so. For example, if an entity encounters a data breach in which the information of 500 or more individuals is compromised, the HITECH Act requires that the entity provide specific details of the breach based upon said protocol [25, p. 2]. In total, the HITECH Act contains four subtitles. Subtitle A is split into two parts: Part 1 covers the improving of healthcare quality, safety and efficiency; Part 2 covers the application and use of health information technology standards and reports [26]. Subtitle B is concerned with testing of healthcare IT and subtitle C covers grants and loans funding for eligible organizations [26]. Subtitle D is also split into two parts and is concerned with privacy and security of electronic health information [26]. Part 1 covers the improvement of the privacy and security of health IT and personal health information in general, part 2 covers the relationship between the HITECH Act and other laws (such as HIPAA which also incorporated the requirements of HITECH into its Final Omnibus Rule in 2013 and thus brought HIPAA and HITECH together into the same legislation) [26]. Additionally, the Office of the National Coordinator (ONC) created the Meaningful Use program with 3 stages to be followed by healthcare organizations adopting EHRs [25, p. 2]. Meaningful use determines the extent to which an entity is utilizing (certified) EHRs in comparison to previous patient documentation methods. [25, p. 2]. Certified EHRs are records that have been certified as meeting defined standards by an authorized testing and certification body [26]. These had to be used in meaningful ways, such as for the exchange of electronic health information to improve quality of care [26]. The program introduced financial incentives that increased every year as new requirements were introduced with each new stage of the program [26]. Failure to follow these requirements would lead to a financial penalty, namely a reduction for reimbursements of Medicare and Medicaid [26]. In order to qualify for federal funds and incentives, facilities had to adopt, but also demonstrate meaningful use of certified EHRs by showing that they had achieved the minimum core objectives in each stage in addition to a set number of menu objectives [26]. It was also necessary to demonstrate compliance with the HIPAA Security and Privacy Rules by conducting risk assessments [26].

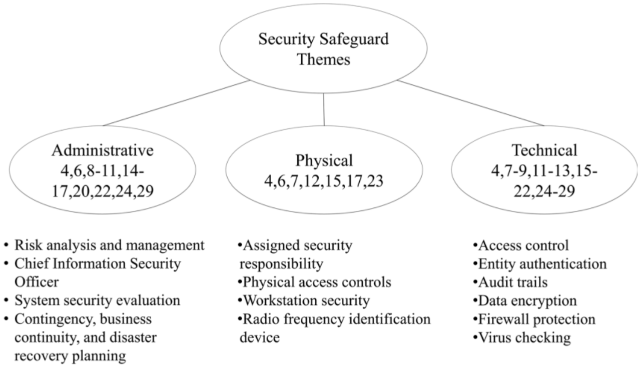

Security Safeguard Themes
Administrative 4,6,8-11,14-17,20,22,24,29
Physical 4,6,7,12,15,17,23
Technical 4,7-9,11-13,15-22,24-29
• Risk analysis and management
• Chief Information Security Officer
• System security evaluation
• Contingency, business continuity, and disaster recovery planning
•Assigned security responsibility
•Physical access controls
•Workstation security
•Radio frequency identification device
•Access control
•Entity authentication
•Audit trails
•Data encryption
•Firewall protection
•Virus checking

Fig. 4. The three themes of security safeguards and their applications according to literature [25, p. 6]

The Patient Protection and Affordable Care Act, also known as PPACA or ACA, was signed on the 23rd of March 2010 and amended on the 1st of May 2010 to the Health Care and Education Reconciliation Act [27]. It provides rights and protections to make health coverage more fair as well as subsidies to make it more affordable in the US [27]. The law is made up of two parts, namely the Healthcare and Education reconciliation Act in addition to the later amended Patient Protection and Affordable Care Act [27]. It generally focuses on improving the affordability and quality of healthcare for Americans (e.g. through increased transparency and reduced discrimination when healthcare providers are choosing who to service/cover), encouraging the development of new patient care models and access to innovative medical therapies as well as setting minimum standards for the services of healthcare providers [27]. The regulation also highlights the role of public programs such as Medicaid and requires improved access to these as well as further improvement in their services and an increased government support to programs such as the Children's health insurance [27]. Part of the regulation also focuses on supporting the healthcare workforce [27]. Another interesting approach from policymakers could also be to implement feedback from stakeholders such as medical experts as they may be able to assist, facilitate and even speed up the process of issuing new suitable legislations or improving existing ones. The Food and Drug Administration (FDA) for example proposed a regulatory framework for AI-enhanced medical devices in April of 2021, and later on in September of 2021 published a new Action Plan which implemented the feedback of stakeholders to its previous regulations [28].

Although laws such as HITECH and similar have encouraged and effectively increased the numbers of healthcare organizations adopting health IT, which brings substantial benefits such as advancing healthcare, improving efficiency and care coordination and facilitating the exchange of health information between different covered entities, it is also important to be aware of the increased likelihood of breaches of patient data due to this increase in the use of electronic records and to enforce additional security mechanisms to ensure healthcare data safety [17, p. 2]. Policymakers should also be aware that some government programs such as the aforementioned Medicare and Medicaid or the Veterans Health and Administration (VHA) might have some specific goals which might cause cybersecurity issues [17, p. 5]. For example, VHA has made significant investments in telehealth over the past few years which might lead to an increase in the number or the diversity of the types of attacks not only to the society and the facilities implementing telehealth, but also potentially to government programs itself [17, p. 5].

Additionally, some experts argue that regulations such as HIPAA limit access to patient data and its exchange even when the data is de-identified as providers fear breaches of privacy [29, p. 387]. A reason for this is that the de-identification method suggested by HIPAA is not completely safe for keeping the patient ID hidden as some information about the patient is still visible which may lead to their identity being exposed. For example, a query to find any patient who is of Indian origin and has some specific cancer diagnosis with a residential zip code 3-digit prefix '479' may result in only one subject; thus exposing the identity of the individual [29, p. 387]. Even without the fear of privacy breaches, sharing of patient data is still a complicated process. For example, "Informatics for Integrating Biology and the Bedside" (i2b2) is a National Institutes of Health (NIH) funded initiative which contains a collection of data systems and over 100 hospitals that are using this software system on top of their clinical database [29, p. 387]. However, in order to be participate, each hospital had to transform their data into a SQL based star schema after de-identification which required much effort [29, p. 387]. This slows or even prevents institutions from exchanging information or making patient data available for research and neither does it facilitate patient participation and involvement regarding their own medical data. All in all, policymakers may need to alter the regulatory environment and frameworks as innovations occur in order to allow the application of these technologies to healthcare [17, p. 5]. For example, some observers believe that blockchain technology offers the possibility of highly secure, decentralized, and longitudinal health records but HIPAA's 1996 security, privacy, and transaction sets are not aligned with blockchain technology [17, p. 5] which might cause an obstacle in its implementation in healthcare. Blockchain technology, especially in context of its possible healthcare applications, is explained in detail in chapter 4.

The European Union (EU) adopted the Data Protection Directive (DPD), officially known as Directive 95/46/EC, in 1995 to protect individuals with regard to collection, processing and the free movement of personal data [30]. The directive is binding for all member countries of the EU so they each have to comply with its regulations. The DPD is built on the 7 principles of "the Organization for Economic Cooperation and Development's Recommendations" (OECD) which were created in 1980 by the "Council Concerning Guidelines Governing the Protection of Privacy and Trans-Border Flows of Personal Data" [30]:

1. Notice: individuals should be notified when their personal data is collected.

2. Purpose: the use of personal data should be limited to the express purpose for which it was collected.

3. Consent: individual consent should be required before personal data is shared with other parties.

4. Security: collected data should be secured against abuse or compromise.

5. Disclosure: data collectors should inform individuals when their personal data is being collected.

6. Access: individuals should have the ability to access their personal data and correct any inaccuracies.

7. Accountability: individuals should have a means to hold data collectors accountable to the previous six principles.

There was, however, no obligation to comply with these principles as they were just a recommendation and there was also a lack of a standard set of rules regarding data privacy throughout Europe [30]. As the European Commission realized that the non-uniform data privacy laws throughout Europe were hindering data flow, they adopted the OECD principles into the DPD which is binding for all EU member states [30]. DPD is also requires that the governing bodies are notified before processing any form of personal data and it is binding even if the data processors are outside of the EU and only using equipment within the EU [30].

The General Data Protection Regulation (GDPR) was later on adopted on the 27th of April 2016 which was enforced starting the 25th of May 2018 with the aim to improve some aspects of the DPD in processing of data and to adapt it to the advancements regarding the digitalization of healthcare systems and data [30]. Today, GDPR is the central data privacy law in the member countries of the EU with some additional national regulations within each country which are required to be compliant with the GDPR [30]. The GDPR reorganizes and updates the DPD's definitions of both "processing" and "personal data" [30]. According to Article 2a of DPD, personal data is defined as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" [31]. Personal data in GDPR is defined in Article 4(1) as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable **natural** person is one who can be identified, directly or indirectly, in particular by reference to **an identifier such as a name, an identification number, location data, an online identifier** or to one or more factors specific to the physical, physiological, **genetic**, mental, economic, cultural or social identity of that natural person" [32] (Differences to DPD in bold). It can be seen that the definition in GDPR was expanded to include specific identifiers such as location data or online identifiers which have become more common since the time when DPD was first introduced. Processing is defined in Article 2b of DPD as "any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction" [31] while it was defined in Article 4-2 of GDPR as "any operation or set of operations which is performed on personal data **or on sets of personal data**, whether or not by automated means, such as collection, recording, organisation, **structuring,** storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction" [32] (Differences in bold). Similarly here it can be seen that not only individual personal data, but also sets of these were considered and that structuring has also been added to the definition of processing. The two mentioned definitions are examples of how the GDPR has made the language used in DPD more strict and improved and expanded DPD to adapt it to the current times.

The Data Protection Directive of EU is similar to HIPAA in the US not only concerning the time in which they were implemented (1995 [31] and 1996 [24, p. 2] respectively), but also that they are the first attempt of each country (or union) to regulate the handling of sensitive personal data and to protect individuals privacy. As it can be seen, both of these regulations were lacking in some aspects such as the strictness of their language, some scopes and with time they also did not cover some newer technologies (such as location information). Thus, they both were improved by newer regulations, namely HITECH in case of HIPAA and GDPR in case of DPD, later on. A big difference between the HIPAA and DPD/GDPR is that the focus of HIPAA lies specifically on the processing of patient health information data [24, p.2] while DPD and GDPR regulate the handling of not health information specifically, but personal information generally [32]. Thus, HIPAA has a more limited scope as it only applies to organizations that handle protected health information while GDPR applies to any type of organization that collects and/or processes any type of personal data. The aforementioned Article 2a of DPD highlights another big difference between the HIPAA (or HITECH) and DPD (or GDPR): The de-identification method suggested by HIPAA is not enough to really de-personalize data according to DPD (as an “identification number” can still be linked to a person as discussed previously) [31]. Therefore, while HIPAA considers de-identification of patient data through an assigned identification number enough for the data to be shared freely, DPD still requires the sharing of such de-identified data to be within its regulatory framework and to follow its laws.

The differences in the rules of each country (or in case of EU a group of countries) make international collaborations considerably more difficult as such projects, for example a project between the US and an EU member country would have to comply with both HIPAA (or rather HITECH) and GDPR which is not standardized yet [12, p. 529] and is difficult to implement in practice. Some of the differences between the regulations of EU members states and those of a non-EU member, in this case the US, were explained above. But the question still remains as to whether there is a difference between regulations and healthcare systems in an EU-member state and a European country which is not part of the EU, but part of the European Economic Area (EEA). In the following, the healthcare IT systems of Austria and Norway will be briefly described and compared to each other and some data privacy regulations of each of these countries and their similarities and differences will also be inspected.

As mentioned before, all EU-members have to comply with the GDPR and Austria is no exception here. Austria currently has a chip-based healthcare system where each insured individual, meaning with an active Austrian insurance contract, is given a chip attached to a physical plastic card. These so called E-Cards were introduced in the year 2005 as a replacement for the then-used paper-based medical certificates (“Krankenschein”) and are nationally used throughout Austria today [33, p. 230]. The card acts as a key in a corresponding Electronic-Management-System, in German “Elektronisches Verwaltungssystem” abbreviated to ELSY, which grants access to the corresponding individual’s medical records and allows the individual to use different insurance services [33, p. 231]. Different facilities such as individual doctor’s ordinations can join and use the system through a so-called O-Card (Ordination Card) [33, p. 231]. After inserting this card into a card reader and entering the correct pin, the ordination is authorized to join the system [33, p. 231]. After authorization, patient information from an E-Card which is inserted into the card reader will be sent to the E-Card system of the ordination from central servers which normally store this information [33, p. 231]. All insured patients and medical facilities or staff get their E-Card and O-Card in a physical letter delivered by post, the PIN for the O-Card is delivered in a separate letter [33, p. 231]. For every visit, after entering an E-Card and checking whether a patient is allowed to have a certain service (such as a check-up), the doctor then books this service and the corresponding cost of this as a bill in the Electronic-Management-System and is reimbursed for this by the patient’s insurance later [33, p. 231]. The patient, if eligible for a service, has to pay nothing or a small portion of the service depending on their insurance and the service. An advantage of this key-based card system is that no data is lost in case the card is damaged or lost as nothing is stored on the chip and it only acts as a key to grant access to the corresponding data [33, p. 231]. E-Cards and their security were recently further improved by adding a picture of the card owner to them as many official documents such as ID-Cards, passports or driver's licenses already include.

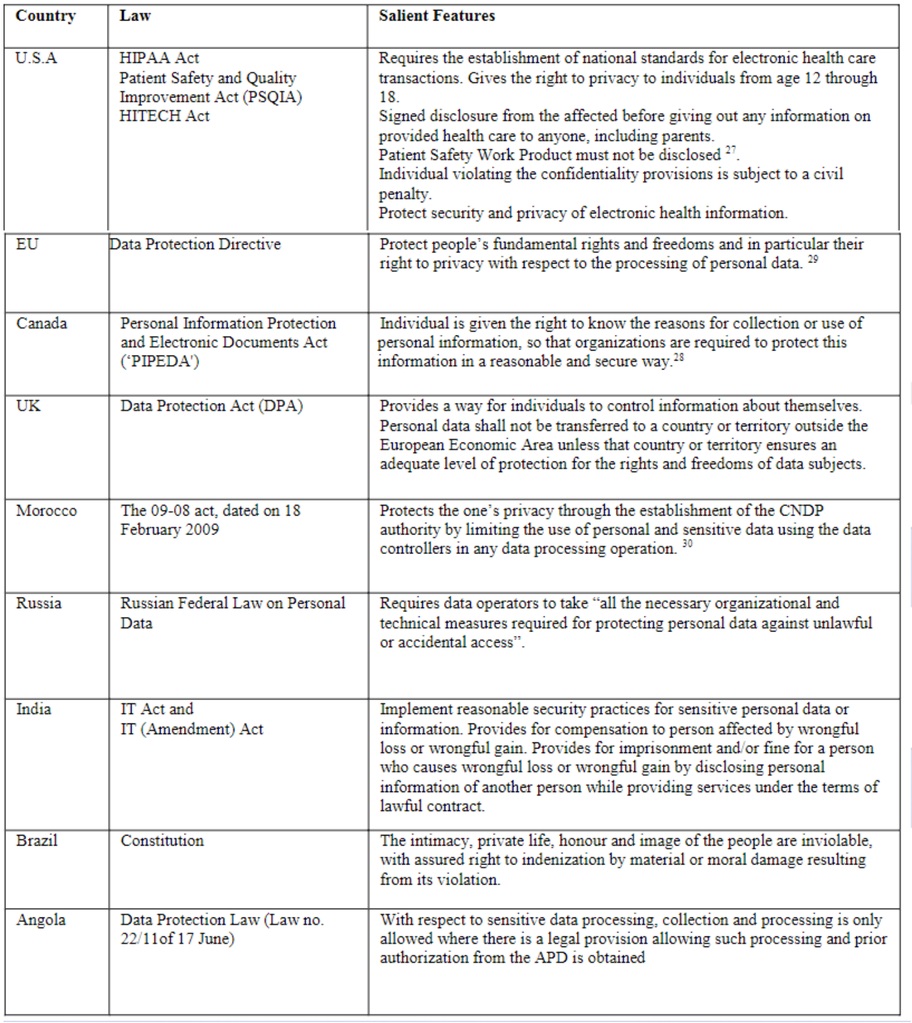
The use of E-Card and generally the social insurance is regulated through the "Allgemeines Sozialversicherungsgesetz" (ASVG), in English "the General Insurance Law", in Austria [33, p. 231]. There are also additional laws for specific cases such as freelance workers (regulated by "Gewerbliches Sozialversicherungsgesetz" or GSVG as well as "Freiberuflich selbständigen-Sozialversicherungsgesetz" or FSVG) or farmers (regulated by "Bauern-Sozialversicherungsgesetz" or BSVG) [34]. Compliance with the General Insurance law is mandatory since the 1st of January 1956 [35]. This law includes information on who has a duty and is required to have health, accident and/or pension insurance in Austria (for example employees from different industries working within Austrian borders), defines the services required of each of the insurance providers as well as the possibilities for voluntary self-insurance for people who are not required to have insurance in Austria [35].

Norway is an example of a non-EU country, it is however (unlike the US) a member of the European Economic Area (EEA). The GDPR applies not only to the EU-members, but also to all countries that are EEA-members and thus, all handling of data either collected from people located in Norway or processed within Norway also have to comply with the GDPR. The GDPR in Norway is implemented in a part of the "Personal Data Act" ("Personopplysningsloven"), from hereon referred to as PDA, which is the central regulation on data protection in Norway and is enforced by the Norwegian Data Protection Authority "Datatilsynet", an independet supervisory authority financed by the Norwegian government [36]. Norway has a national online healthcare system "Helsenorge" (directly translates to "Health Norway") with an individual so-called "Pasientjournal" ("Patient journal") for each insured patient, where the doctors or physicians write down visits and different diagnostics as well as the patient's medical history and previous sicknesses [37]. In case a medicine is prescribed or a referral to a specialist is needed, the general practitioner doctor will create an entry in an online list which is stored in the patients profile. The patient profile is linked to their Norwegian national security number, so they only need to share this number with the specialist or a pharmacy and the staff there will then have access to their online receipt or referral on the patient's profile after entering this number into their authorized computers and systems and can provide the needed services or medicine for the patient [37]. The Norwegian national security number is similar to the Austrian health insurance number but it is used more widely and in several areas such as healthcare, banking and for taxing and employment purposes in Norway (and not only for health insurance like in

Austria). This system eliminates any need for paper and is completely internet based. There are recently some projects, such as the new "Helseplattformen" in mid-Norway, in motion to standardize the entries in the patient journal as their format currently can vary from doctor to doctor and some entries are redundant and generally make a good overview more difficult [38]. The handling of all patient data on these journals and generally in a patient's online profile falls under the GDPR (or Norway's adaptation of this regulation).

The Personal Data Act in Norway is mostly identical to the GDPR in different areas such as the scope of application, key definitions (of for example personal data) and legal bases with some minor adjustments or more detailed specifications where GDPR allows it [36]. An interesting addition is that in Norway, the Norwegian King can decide if the law does not apply to some institutions or in certain situations according to §2 of the PDA [39]. Article 8(1) of the GDPR generally specifies the minimum age of children, whose consent is needed to process their personal to be 16 years, if the children are younger than 16 years old, then the consent of the person with the parental responsibility for them is needed instead [32]. According to the §5 of the PDA however, the consent of the children (and not their guardians) is required if they are 13 years old or older [39]. The GDPR was added to the EEA contract on the 25th of May 2018 and was implemented in the PDA in Norway on the 15 of June 2018, and was enforced officially starting the 20th of July 2018 [36]. This new regulation, namely PDA ("Lov 15. juni 2018 om behandling av personopplysninger")[39] replaced the existing Personal Data Act ("Lov 14. April 2000 om behandling av personopplysninger") which was made on the 14th of April 2000 and was enforced from 1st of January 2001 [40] until it was replaced by the PDA in 2018.

In this section, specific laws concerning the improvement of healthcare IT and the protection of systems and data of medical facilities and other personal data in general were discussed. The comparison of the healthcare systems and respective regulations of Norway, a European but non-EU member, and Austria, an EU-member was in the center of the final part of the chapter. There was also a large focus on different regulations and the development of the data privacy laws especially concerning electronic health records in the US as an example of a non European and non-EU country for a general comparison's sake. The points in this chapter do not apply to all systems and situations and naturally cannot be generalized for all countries whether they are in Europe or an EU-member or not. To provide further context and a brief more general overview for some additional countries, figure 5 provides a list of data privacy laws of a broader, more representative group of countries from around the world. It is interesting to see that for example Brazil only has a very general regulation in it's constitution to protect "people's honor" and does not enforce any additional data privacy regulations (for health-related data or otherwise).



Country	Law	Salient Features
U.S.A	HIPAA Act Patient Safety and Quality Improvement Act (PSQIA) HITECH Act	Requires the establishment of national standards for electronic health care transactions. Gives the right to privacy to individuals from age 12 through 18. Signed disclosure from the affected before giving out any information on provided health care to anyone, including parents. Patient Safety Work Product must not be disclosed [27]. Individual violating the confidentiality provisions is subject to a civil penalty. Protect security and privacy of electronic health information.
EU	Data Protection Directive	Protect people's fundamental rights and freedoms and in particular their right to privacy with respect to the processing of personal data. [29]
Canada	Personal Information Protection and Electronic Documents Act ('PIPEDA')	Individual is given the right to know the reasons for collection or use of personal information, so that organizations are required to protect this information in a reasonable and secure way.[28]
UK	Data Protection Act (DPA)	Provides a way for individuals to control information about themselves. Personal data shall not be transferred to a country or territory outside the European Economic Area unless that country or territory ensures an adequate level of protection for the rights and freedoms of data subjects.
Morocco	The 09-08 act, dated on 18 February 2009	Protects the one's privacy through the establishment of the CNDP authority by limiting the use of personal and sensitive data using the data controllers in any data processing operation. [30]
Russia	Russian Federal Law on Personal Data	Requires data operators to take "all the necessary organizational and technical measures required for protecting personal data against unlawful or accidental access".
India	IT Act and IT (Amendment) Act	Implement reasonable security practices for sensitive personal data or information. Provides for compensation to person affected by wrongful loss or wrongful gain. Provides for imprisonment and/or fine for a person who causes wrongful loss or wrongful gain by disclosing personal information of another person while providing services under the terms of lawful contract.
Brazil	Constitution	The intimacy, private life, honour and image of the people are inviolable, with assured right to indenization by material or moral damage resulting from its violation.
Angola	Data Protection Law (Law no. 22/11of 17 June)	With respect to sensitive data processing, collection and processing is only allowed where there is a legal provision allowing such processing and prior authorization from the APD is obtained

Fig. 5. A short overview of data protection laws in some countries [21, p. 76-77]

3 Resilience

Resilience is a central concept when discussing the ensuring of cybersecurity and the security of data within an organization. In the previous chapter, some of the common cybersecurity measures currently used in organizations and facilities (especially in healthcare) were discussed. Some of the weaknesses of these measures, such as guessing very simple or common passwords, were also highlighted in chapter 2.2. These weaknesses can indeed be reduced by combining several technologies or through raising awareness level of the employees and the personnel, they can however never be fully overcome so it is not possible to protect a system or an organization from all cyberattacks forever and it will very likely be compromised at some point. In this chapter, we will discuss a central concept, namely resilience, for strengthening a system and reducing the likelihood of successful attacks especially under stress situations (such as a pandemic). Additionally, a resilient system ensures that the damage and loss is kept to a minimum even in case of a successful attack and data breach and that the system can resume its working (almost) normally during or quickly after the damage. There will be a short definition of this concept as well as how organizations can become more resilient to minimize damage and data loss and recover more quickly from cyberattacks.

Resilience is defined as the ability of an organization to withstand disruptions to its operation and to recover from them. Resilience in the context of organizations is referred to as operational resilience from here on out. These disruptions to an organization's operations can generally be man-made or natural, here we will focus on man-made disruptions, specifically malicious cyberattacks. The definitions for operational resilience vary in different works of literature, however, these perspectives can be divided into three general groups: Input-base resilience (IBR), output-base resilience (OBR) or a combination of the two [41, p. 2]. The IBR perspective defines a system's ability to respond and recover from disruptions based on its score on IBR scales, which is determined by indicators such as the system's flexibility, buffers, visibility, disruption preparedness, agility, collaboration, integration and information sharing amongst others [41, p. 2]. The OBR perspective on the other hand states that a system's resilience level cannot be determined before a disruption has happened [41, p. 2]. The core elements in this perspective are the system's disruption absorption, recoverability, adaptability and transformability [41, p. 2]. Several studies with this perspective argue that IBR elements do not necessarily imply variability in resilience unless they are bundled and considered together in some cases which requires a high amount of work and may not be easily possible in practice [41, p. 2-3].

Operational resilience consists of two main dimensions which are also of importance within the aforementioned OBR perspective: disruption absorption and recoverability [41, p. 1]. The disruption absorption dimension is defined as the ability of a system to maintain the structure and normal functioning of operations when faced with disruptions, and the recoverability dimension is defined as the ability of a system to restore operations to a prior normal level of performance after being disrupted [41, p. 3]. Knowing the normal operating performance level before the occurrence of a disruption, operational resilience can be determined by [41, p. 3]:

1. calculating the magnitude of the drop in normal operating performance level immediately after the occurrence of a disruptive event (and before the start of a recovery action).
2. calculating the time it takes to restore operations to normal performance level after the start of the recovery action.

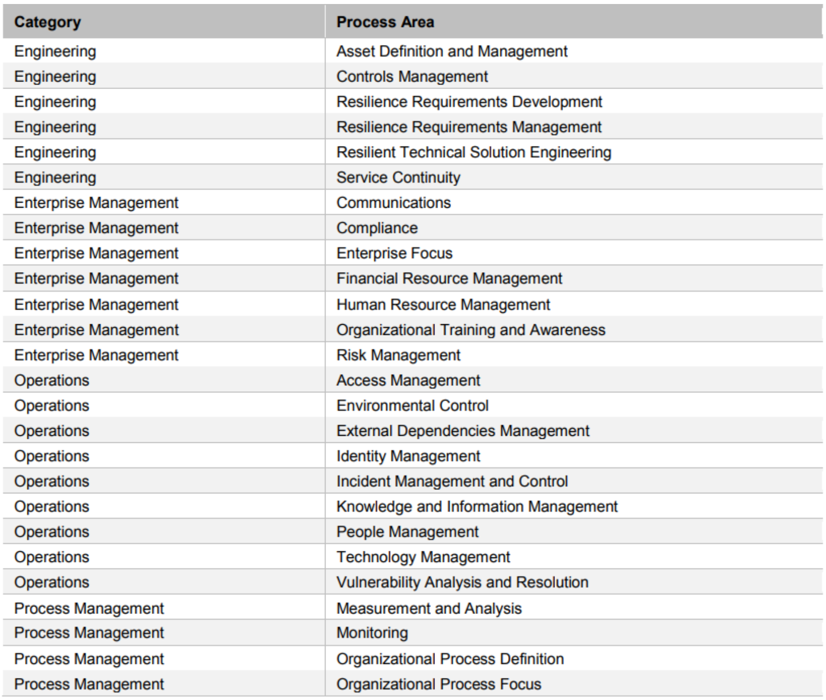
A smaller drop in the performance level suggests that the system possesses disruption absorption capability while a greater drop in the performance level suggests that the system does not posses this capability. The recoverability capability of the system can be measured based on the recovery time, where a longer recovery time suggests that the system lacks recoverability while a shorter recovery time shows that the system possesses recoverability capability. It is important to note that while these two capabilities, namely disruption absorption and recoverability, complement each other, there is no causation between the two dimensions [41, p. 3]. This means that the fact that a system or organization has disruption absorption capabilities does not automatically imply that it has recoverability capabilities or vice versa, and therefore both of these aspects should be considered and implemented individually in the system to ensure the best possible level of resilience.

The CERT Resilience Management Model (CERT-RMM), first developed by researchers from Carnegie University, is an example of an approach which aims to support and improve the management of operational resilience in complex and uncertain business environments [42]. Security and continuity or resilience are often only considered from a technological perspective, meaning that organizations often seek technological advancements to keep their data secure and to improve the way they handle disruptions and incidents. While technological security measures such as the ones mentioned in the previous chapter are indeed essential to cybersecurity, they are not the only aspect which should be considered here. Incidents such as security breaches can often be traced back to poorly designed and managed processes at the enterprise and operational levels, not technology failures [42, p. 5]. Therefore, especially in the context of operational resilience, it is important to have clearly defined processes and comprehensible process management in order to be able to successfully and efficiently manage an organization's existing technologies [42, p. 5].

An important concept for the CERT-RMM is convergence which is the idea of harmonization of operational risk management activities that have similar objectives [42, p. 17]. Many organizations are now beginning to realize that security, business continuity, and IT operations management are complementary functions that are all focused on managing operational risk and have the same goal, namely to improve and sustain operational resilience [42, p. 17-18]. This view is often supported by collaborative practice codes in each domain such as security practices, which reference business continuity and IT operations management practices and thus acknowledge that security practices alone do not address the conditions and consequences of risks, becoming more popular and widely used today [42, p. 18]. It is efficient to use a common, collaborative approach when organizational functions share many of the same objectives, issues and solutions [42, p. 18]. Generally, the extent to which convergence has been achieved directly affects the level of operational resilience, which in turn affects the ability of an organization to meet its goals and mission [42, p. 18]. Security planning and management, business continuity and disaster recovery management, and IT operations and service delivery management are bound by the same operational risk drivers and therefore are likely to have risks in common that can be managed using similar (or identical) approaches, thus eliminating redundancy which in turn reduces costs [42, p. 18]. These functions have been separated in traditional organizational structures and have a long history of working independently [42, p. 18] which is one of the reasons why convergence is not fully achieved in many organizations today and the process management aspect is overlooked when discussing security measurements. This has, however, been improving which is reflected in collaborative codes of practice gaining popularity in the recent years as mentioned above.

The CERT-RMM allows organizations to identify the current level of organizational capability, set an appropriate and realistic desired performance target, measure the gap between current and targeted performance, and develop action plans to close this gap by using a process definition as a benchmark [42, p. 2]. The CERT-RMM is the first known model in the security and continuity domain that includes a capability dimension, meaning it allows an organization to measure its ability to control operation resilience and predict how it will perform under disruption [42, p. 2]. The CERT-RMM includes a detailed evaluation of process areas throughout an organization. A process area in this model is defined as "a cluster of related practices in an area that, when implemented collectively, satisfy a set of goals considered important for making improvement in that area" [42, p. 31]. The first version of CERT-RMM contains 26 process areas which cover four general areas of operational resilience management, namely Engineering, Enterprise Management, Operations, and Process Management [42, p. 31]. A list of all these process areas and their corresponding management area category can be seen in figure 6. All in all, CERT-RMM aids in determining, implementing and managing activities which help ensure resilience of services and thus operational resilience in individual organizations [42, p. 9]. This model, although originally constructed in the financial services industry, was developed to be scalable across different organizations regardless of their industry or size and is already being introduced and used in other small and large industrial sectors and government organizations [42, p. 9-10].

Some of the aforementioned cybersecurity measures such as creating backups also contribute to resilience which underlines the aforementioned convergence



Category	Process Area
Engineering	Asset Definition and Management
Engineering	Controls Management
Engineering	Resilience Requirements Development
Engineering	Resilience Requirements Management
Engineering	Resilient Technical Solution Engineering
Engineering	Service Continuity
Enterprise Management	Communications
Enterprise Management	Compliance
Enterprise Management	Enterprise Focus
Enterprise Management	Financial Resource Management
Enterprise Management	Human Resource Management
Enterprise Management	Organizational Training and Awareness
Enterprise Management	Risk Management
Operations	Access Management
Operations	Environmental Control
Operations	External Dependencies Management
Operations	Identity Management
Operations	Incident Management and Control
Operations	Knowledge and Information Management
Operations	People Management
Operations	Technology Management
Operations	Vulnerability Analysis and Resolution
Process Management	Measurement and Analysis
Process Management	Monitoring
Process Management	Organizational Process Definition
Process Management	Organizational Process Focus

Fig. 6. An overview of CERT-RMM's Process Areas and their corresponding category [42, p. 31]

between these areas. Storing several copies of a file in different locations, by for example using cloud services, and thus ensuring the availability of the file through its other copies in case one of the copies in a location is compromised is also considered a resilience strategy. This is because the file's availability in this case allows continued access to it despite one version of it being compromised and thus the system can continue its operation as normal which corresponds to the above definition of resilience. As it can be seen, resilience is closely related to some other cybersecurity measures and concepts and thus once again, combining technologies or implementing appropriate measures in one area (e.g. availability) may also improve the other (e.g. resilience).

In the context of healthcare, the focus on how practices need to cope, respond and adapt to stress has increased in healthcare studies in this field in the recent years [43, p. 1-2]. This focus on resilience brings a new perspective into these studies and provides a connection between different interests (e.g. agendas or strategies) across different healthcare levels and contexts which contributes to a better understanding of complex healthcare systems [43, p. 2]. The majority of research on resilience in healthcare has focused on direct disruptions and crises such as natural disasters or pandemics, however, the attention to the importance of resilience in everyday healthcare operations has increased recently [43, p. 2]. With a focus on resilience in everyday operations, resilience in healthcare can be defined, similar to before, as the capacity to adapt to challenges and changes at different system levels in order to maintain high quality care [43, p. 2]. A system's ability to continue operation under continuous stress can also strengthen its ability to manage well in situations with sudden disruptions and shocks, therefore further supporting and justifying the focus on everyday resilience [43, p. 2].

There is a range of research aiming to provide methods to operationalize and implement resilience in healthcare. There is some ambiguity and differences in how different researchers define key characteristics or markers of resilience in healthcare. As mentioned in both the IBR and OBR perspectives, one way to empirically measure a concept in practice is to establish indicators which in this case, however, is not straightforward as seen in the aforementioned ambiguity of researches in this field. There can be no single indicator, but rather a bundle of interrelated factors for resilience due to its complex nature and its various relating factors and processes which makes providing a clear definition of the involved variables and concepts challenging [43, p. 2]. Developing these indicators and establishing suitable bundles of interrelated factors requires an extensive amount of work as mentioned previously, it is nonetheless important as it allows organizations to identify and understand their strengths and weaknesses and thus better prepare for and respond to stress and challenges [43, p. 2]. A concept that occurs in most studies is the importance of adaptive capability, followed by concepts such as leadership and awareness, planning and anticipation which are also mentioned in most studies [43, p. 2]. The Concepts for Applying Resilience Engineering (CARE) model is a framework for understanding the importance of adaptive capacity which is aims to illustrate a simplified abstraction of the key resilience concepts to be investigate empirically as well as the relationships between them in complex healthcare systems [44, p. 3]. As show in figure 7, the CARE model consists of three main concepts, namely Work as Imagined (WAI), Work as Done (WAD) and Outcomes. In this model, WAI is defined as an intended or imagined alignment between the demands in system (e.g. service quality) and the system's capacity to meet those demands [44, p. 3]. In practice, demand and capacity will never be aligned due to the complexity of the system and factors such as unforeseen disruptions and therefore adaptations and adjustments have to be carried out [44, p. 3]. These adjustments to overcome and accommodate the misalignment between the demand and capacity as well as natural variances in how tasks are carried out in practice are conceptualized as WAD [44, p. 3]. Outcomes in this model, such as consequences for patients or staff, are defined broadly where success and failure are not fixed categories and are subject to interpretation based on the individual organization's judgment or in the context of the task [44, p. 3].

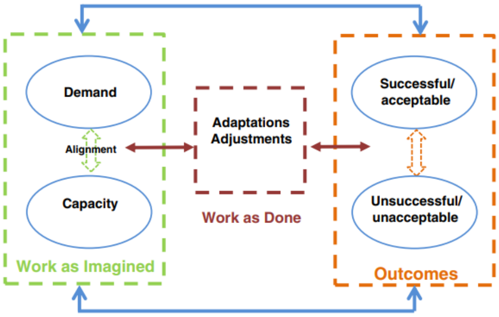


Fig. 7. The Concepts for Applying Resilience Engineering (CARE) model [44, p. 3]

A study by Lyng et al. extracted 10 organizational resilience capacities in healthcare as a result of analyzing 25 different research projects from a range of empirical healthcare settings (e.g. hospitals, nursing homes) and also interviewing 16 researchers involved in those projects [43, p. 4]. An overview of these organizational resilience capacities as well as their sub-themes can be seen in figure 8. Structure here refers to structures that support work and practice within organization, and it is made up of four sub themes: Technology (accessibility and compatibility of different software and technology), roles and responsibilities (stability among staff and clearly determined responsibilities), arenas (meeting arenas for face-to-face communication and learning) and plans (plans and procedures of healthcare practices) [43, p. 6]. Learning describes how the organization facilitates and provides learning activities and opportunities and includes: Collaborative learning (learning through interactions between stakeholders) which leads to knowledge acquisition and training [43, p. 7]. Alignment refers to adaptions to manage what is required at any given time and its circumstances, and is made up of: Adapting (of practices and care to specific patient needs), aligning (establishing shared goals and understanding) and self-organizing (of healthcare personnel) [43, p. 7]. Coordination refers to how work and information flow across different disciplines and other organizations is organized and includes: Care coordination, collaboration, buffers and continuity (of staff, resources, and learning) [43, p. 7]. Leadership concerns how leaders support, motivate and contribute to the organization based on: Leadership interaction (staff support and motivation), prioritizing (between conflicting demands and capacities), inclusion and empowerment [43, p. 8]. Risk awareness refers to how the organization understands and reflects on risks that may affect the patient and possible adverse events and includes: Proactive responses, reactive responses and risk perception [43, p. 8]. Involvement refers to how the organization introduces and involves different healthcare system actors, and whether it systematically informs itself of ongoing situations involving: Family, patients and other stakeholders [43, p. 9]. Competence is defined as having the appropriate knowledge, attitude, skills, and experience for decision-making, being able to take on necessary adaptations, and to have the situational understanding needed to provide quality care [43, p. 9]. Facilitators concerns how the organization or employees ease positive impacts for the organization [43, p. 9]. There are two facilitator roles, namely knowledge brokers who facilitate knowledge transfer among colleagues and across boundaries, and champions who facilitate through their own actions [43, p. 9]. Lastly, communication includes the sub-themes translating (of information to the specific receiver) and communicating (having an awareness of the amount and type of information to be transferred depending on the situation and an openness for feedback) [43, p. 10].

4 Data Integrity through Blockchain

Data integrity is the second important aspect of ensuring the security of an organization and its data in addition to resilience. In this chapter, the concept of data integrity will be defined and some possible implementations of this concept through the use of blockchain technology will be discussed. Blockchain technology as well as how it operates and a variation of it, namely the permissioned blockchain, which is more suitable for use with medical health records due to its increased confidentiality, will be introduced in the first sub-chapter. In the second sub-chapter, Zero-knowledge proofs are introduced as a possible alternative to permissioned blockchains which can be used in addition to the original concept of (public) blockchains for different purposes such as increasing confidentiality and restricting access to data (e.g. through an identity claim model mentioned at the end of the sub-chapter) amongst others. This is made possible as zero-knowledge proofs allow the data on blockchain to be encrypted (not public anymore) and yet still be verifiable as needed. The encryption of data, as opposed to having public access, significantly reduces security risks such as possible revelations of the identities of the parties involved in a transaction and

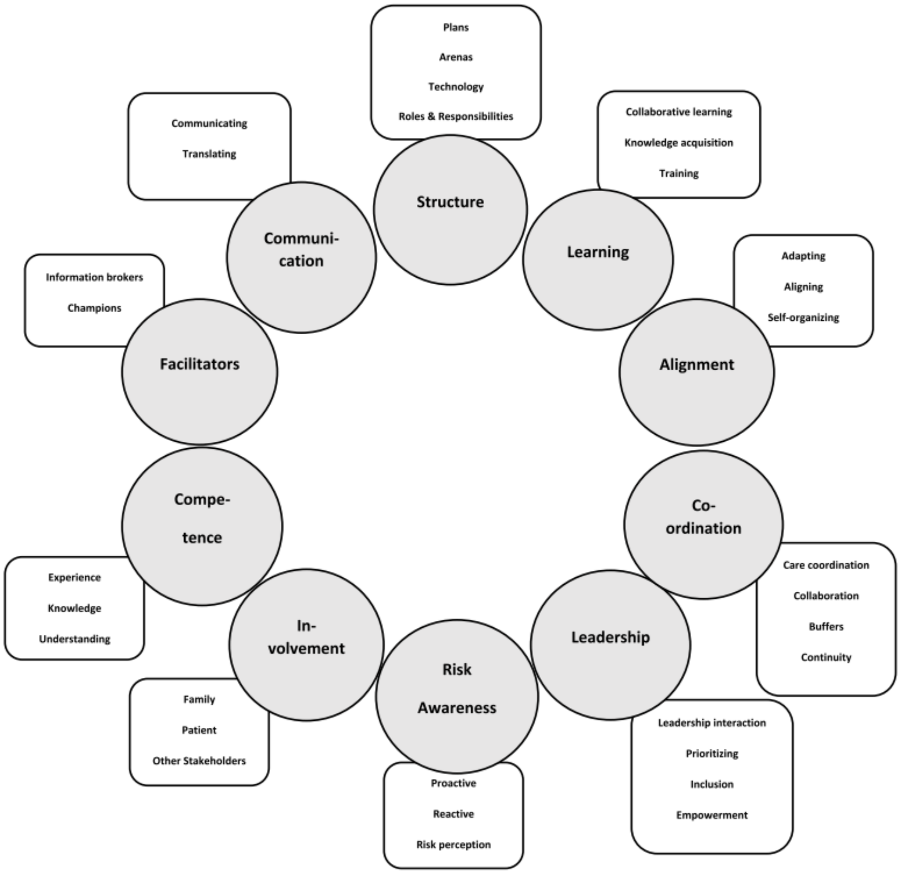


Fig. 8. Resilience capacities with associated sub-themes, ordered clockwise based on occurrence frequency ("Structure" with the most identified instances and "Communication" with the least) [43, p. 5]

other information falling into the wrong hands of malicious parties. Finally, the current legal situation and different laws and regulations regarding blockchain in a selection of countries is discussed in the third and final sub-chapter.

4.1 Blockchain and Data Integrity

Data resilience and data integrity are both essential to ensure the safety and accuracy of data. In case of a successful attack and overcoming other security measures such as access control policies, the attacker will have access to data which already causes damage and problems by for example revealing the identity of patients and their health status (in case of medical records) which may be misused for blackmail or otherwise by the attacker. In order to minimize or prevent any further damage such as data loss, it is important that the organization has implemented resiliency measures as mentioned previously. Implementing measures to guarantee data integrity will also significantly reduce the damage and loss of data (or data being replaced by incorrect data) by preventing the attacker from modifying the data they have access to after a successful attack. Data resilience was defined in the aforementioned section in detail, data integrity will be defined in the following.

We previously defined access control mechanisms as a cybersecurity measure currently in use in many facilities, medical or otherwise, to prevent unauthorized access to services and data. Some possible life-endangering consequences of unauthorized parties gathering access to sensitive data such as health records were also discussed. Data integrity will in this work be defined as preventing unauthorized modification of data [45, p. 291]. So one can deduce that the ability for unauthorized parties to not only access sensitive data, but to also modify them, especially without the knowledge of responsible authorities, can only lead to more severe and life-threatening or fatal consequences and endanger many patients in case of medical records. Unauthorized access may, but does not always, lead to the disturbance of data integrity. There is an increased need for models and mechanisms to prevent unauthorized manipulation or modification of data with the increasing number and complexity of the data resources of most organizations.

Integrity was defined in the above in relation to data, and their security and privacy, as preventing unauthorized modification of data. There are, however, several different definitions proposed in the literature for the general concept of integrity which shall be included here for the sake of completeness. Zviran & Glazer [45, p. 293] divide these definitions into three major groups: single-element data-focused definitions, single-element non-data-focused definitions, and multi-element focused (comprehensive) definitions. These definitions focus either on a single element or a group of elements [45, p. 293]. According to the data-focused definition, data integrity is concerned with the correctness of the database content which can be compromised by failures caused by the (intentional or unintentional) actions of users, programs or systems [45, p. 293-294]. Thus, data integrity is viewed in this group as a complement to data security consisting of semantic integrity, concurrency control and recovery mechanisms [45, p. 293-294]. Most works of literature in this group suggest and define constraints which may for example prevent the modifications of data or trigger automatic recovery mechanisms if violated [45, p. 293]. As the name suggests, non-data-focused definition focuses more on the integrity of systems instead of data [45, p. 294]. A system is said to possess integrity if it adheres to a well-defined code of behavior and performs as it was intended by its creators [45, p. 294]. It it said to be a property of state, in which a machine or system is correct overall if and only if all of its states are correct[45, p. 294].

A multi-element focused definition mentioned by Zviran & Glazer [45, p. 294] covers the following areas:

1. How correct the information is thought to be
2. Level of confidence that the information is from the original source.
3. Correctness of the functioning of the process using the information.
4. Level of correspondence of the process function to the designed intent.
5. How correct the information in an object is initially.
6. Confidence that the information in an object is unaltered.

The Integrity Working Group (IWG) also developed a definition where data integrity is referred to as “a property that data, an information process, computer equipment and/or any software, people, etc. or any collection of these entities meet an a priori expectation of quality that is satisfactory and adequate in some circumstances. The attributes of quality can be general in nature and implied by the context of the discussion, or specific and in terms of some intended usage or application” [45, p. 294]. The definitions in this third group seem to be the most comprehensible and extensive as they cover not only data or only systems, but both in addition to several other elements such as software as well as several areas such as the source of information.

The increased threat to data integrity leads to increased concerns especially in case of medical facilities. Unauthorized modification of health records can lead to a patient for example receiving the wrong medicine which may be fatal [46, p. 40613]. It is therefore very important to prevent this at all costs. Some measures such as the common cybersecurity measures mentioned in previous chapters, e.g. authentication, encryption and securing clouds, are suggested as a way of preventing successful attacks and access to data [46, p. 40616]. Preventing unauthorized access to data also prevents their modification as one cannot modify data without having access to them in the first place. Many researchers propose the use of Blockchain and provide different architecture and/or possible approaches for the implementation and use of this technology for data integrity in healthcare facilities [46, p. 40616].

Blockchain, also known as Distributed Ledger Technology (DLT), was first introduced in 2008 in a paper by an anonymous individual or group of individuals known as “Satoshi Nakamoto” [47]. The paper introduced this technology in the form of Bitcoin, a decentralized peer-to-peer electronic cash system [47, p. 1], and the ideas and models used in this paper are used to this date for Bitcoin as well as many other cryptocurrencies. The main idea here is to eliminate the need for a trusted third party, for example financial institutions, in online electronic transactions through the use of a peer-to-peer(P2P) network and providing a cryptographic proof instead of trust [47, p. 1]. Most business applications today use centralized networks where the data is stored on a central system and the communications and modifications of data are controlled by a central authority, namely a server. The so-called clients or users of the system communicate by sending requests to the server to which the server responds. Here, all information is stored on one server which also is responsible for all client requests [48, p. 102] and a failure of the server can lead to malfunctioning and information loss . The complexity of these systems also increases with size and they are often not easily expandable and scalable without suffering service quality losses. A P2P network attempts to address these problems using a decentralized approach where each node is not either a client or a server, but a so-called "servent" (combination of the words "server" and "client") and so can act as both [48, p. 101]. Schollmeier gives a more precise definition of this concept saying that "A distributed network architecture may be called a Peer-to-Peer (P-to-P, P2P, ...) network, if the participants share a part of their own hardware resources (processing power, storage capacity, network link capacity, printers, ...) [48, p. 101]. These shared resources are necessary to provide the Service and content offered by the network (e.g. file sharing or shared workspaces for collaboration): They are accessible by other peers directly, without passing intermediary entities. The participants of such a network are thus resource (Service and content) providers as well as resource (Service and content) requestors (Servent-concept)" [48, p. 101].

As the name suggests, a blockchain is a collection of so-called blocks, with each block containing one or several records of transactions. To add a new (set of) transaction(s) to the block, first the new transactions are broadcast to all nodes in the corresponding network [47, p. 3]. Each node then collects the transaction(s) into a block and finds, in case of Bitcoin, a proof-of-work (or other cryptographic algorithms for calculating a difficult hash-value and verification) and broadcasts this back to all peers/nodes in the network [47, p. 3]. Nodes then accept a block if all transactions in this block are valid and not already spent by verifying the hash-value also known as its signature [47, p. 3]. Blockchain thus solves the double-spending problem, meaning that it ensures the owner of an electronic coin did not spend the same coin twice or more in several transactions, without the need for a trusted central authority to check this and instead through consensus mechanisms [47, p. 2]. There are different consensus mechanisms, such as Proof of Work mentioned before, which all conform to a democratic decision-making process. If a block is accepted by a node, it will create the next block in the chain using the hash-value of the accepted block while creating the hash-value of the next block [47, p. 3]. The hash-value of the blocks guarantees the immutability, meaning that no transaction can be modified or reversed once accepted and appended to the blockchain. Changing a transaction would require recomputing the signatures/hash-values of that block and all the consequent blocks [47, p. 3]. This is a very tedious process and almost impossible due to requiring high amounts of resources amongst others. Additionally, each transaction (on a block) is verified by several nodes and as long as the majority of the nodes remain uncorrupted, disturbing the hash-values will be detected by the nodes in the network and the changes will be deemed invalid [47, p. 3]. Thus, data integrity is guaranteed even without a central authority. Additionally, every peer in the network can see all data in the block and all blockchains and the (amount and time of the) transactions included in them are visible and freely accessible for the public (without disclosing the identity of the parties in the transaction) which ensures transparency [47, p. 6]. Blockchain also allows interoperability due to its distributed network structure. In medical context for example, the equipment of different care providers can be registered as members of the blockchain and the transfer of patient information among these members can then be implemented securely and efficiently using smart contracts. Smart contracts are agreed-upon conditions which are automatically executed and are trusted by all members of a blockchain.

Some common consensus mechanisms which may be used in the above process to sign and verify blocks are Proof of Work (POW), Ripple Protocol Consensus Algorithm (RPCA) and Proof of Stake (POS), which are currently in use by the cryptocurrencies with the highest market cap [49, p. 1546]. POW is also used (and was first introduced) in Bitcoin and works by scanning for a value that, when hashed, has a hash which begins with a number of zero bits [47, p. 3]. This is accomplished by adding a so-called "nonce" (a disposable number used once) to the original value and incrementing it until the block's hash has the required zero bits [47, p. 3]. Once this has been achieved, the proof of work is satisfied and the block cannot be changed without redoing all the blocks after it [49, p. 1546]. As shown in figure 9 each block in the chain has a hash which consists of the previous block's hash and the corresponding nonce. The first block in a chain is an exception as it has no previous block so its hash is entirely zeroes [49, p. 1546]. The longer a blockchain is (higher number of linked blocks), the more difficult it will be to manipulate and the more secure it will be. RPCA is used exclusively by the Ripple cryptocurrency and addresses latency issues present within other consensus mechanisms [49, p. 1546]. Here, each server creates a public list called "candidate set" which includes all valid transactions that server has access to, and then combines its own candidate set with that of all other servers it has kept a reference to (this list of other servers is known as the server's "unique node list") [49, p. 1546]. Afterwards, each server votes on the validity of each transaction over one or multiple rounds of voting and finally all transactions which are accepted by at least 80% of the servers are written to the public ledger which is then closed and cannot be changed afterwards [49, p. 1546]. POS was originally implemented in 2012 in the cryptocurrency PeerCoin, where it uses POW for the initial coin minting and POS mainly for network security [49, p. 1546]. POS uses the concept of "coin days" to establish the age of the coins used in a transaction; for example holding 10 coins for 10 days returns a coin-day value of 100, which is reset to zero after spending these coins in a transaction as here the age of coins is consumed [49, p. 1546]. In order to add a new transaction or block to the network, a validator has to pay themselves through consuming their aforementioned coin age [49, p. 1547]. In POS, the chain with the highest consumed coin age is considered to be the main chain unlike POW where the chain with the highest number of transactions is seen as the main chain [49, p. 1546].

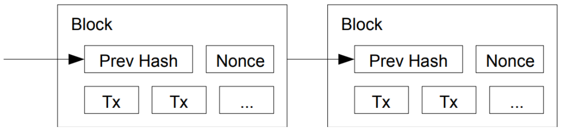


Fig. 9. Visualization of two blocks in a Proof-of-Work blockchain [47, p. 3]

Blockchain guarantees data integrity, meaning that data cannot be changed by unauthorized parties without notice [45, p. 291], however there is still the possibility of unauthorized access and reading of data. A solution for this would be to use Blockchain in combination with the aforementioned technologies such as authentication or access control, which limit or prevent unauthorized access to data [46, p. 40616]. There are already some works of literature which introduce approaches and models to integrate blockchain into existing cybersecurity measures. The article by Di-Francesco-Maesa et al. [50, p. 94] proposes an integration of blockchain into existing attribute-based access control systems by essentially codifying these access control policies into the aforementioned smart contracts and deploying, storing and executing them on a blockchain and thus implementing these policies in a decentralized self-evaluating manner. A different work by Liu et al. [51, p. 468] proposes a Data-Integrity-as-a-Service (DIaaS) platform where blockchain is used to ensure data integrity for cloud-based IoT applications.

As it can be seen, the combination of blockchain with existing cybersecurity measures can increase the security overall. To further increase the security of blockchain, it is possible to combine this with the zero-knowledge proof mentioned in the next sub chapter 4.2. However, there are also downsides, such as potential scalability issues, when using Blockchain in large distributed systems such as smart healthcare [52, p. 2]. Additionally, the public access to data stored on a blockchain is not suitable for using with sensitive health records [52, p. 2]. To address these issues, a more confidential and scalable variation of the current Blockchain structure called “Permissioned Blockchain” is being developed [52, p. 2]. There are two types of permissioned blockchains: private and consortium blockchains, both of which run on a private network and direct access to their data and submitting transactions is restricted to a predefined set of entities or parties [52, p. 2]. In private blockchains, write permissions (to modify data) are centralized and kept to one entity (e.g. an organization) while the read permissions (to access and read data) may be public or restricted to some extent the latter of which offers increased privacy [52, p. 2]. This means that only the one authorized entity needs to approve a transaction once it is created (only possible for authorized nodes to create new transactions) in order for the transaction to be added to the blockchain without the need for any cryptographic hashing which reduces overall costs [52, p. 3].

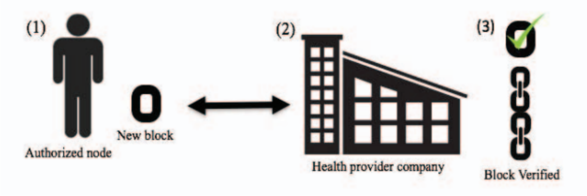


Fig. 10. Example of the working of a private blockchain in healthcare [52, p. 2]

Figure 10 illustrates how a private blockchain might operate in healthcare: in step 1, a new block (or transaction) is generated by an authorized node, which is then approved by the authorized health provider company (with write permissions) in step 2 and finally added by that company to the existing blockchain in step 3 [52, p. 3]. In consortium blockchains, a new block or transaction has to be approved by the majority of a pre-selected set of trusted nodes in order to be added to the existing chain or ledger; the right to read this chain may be public (similarly to private blockchains) or made restricted to only participants (namely the pre-selected set of trusted nodes) [52, p. 3]. Figure 11 visualizes how a new block created by an authorized node such as a doctor in step 1 may be added to a consortium blockchain (in step 3) only after the majority of the pre-selected set of trusted nodes (in this case for example 3 out of 5) have verified it in step 2 [52, p. 3]. Unlike private blockchains, consortium blockchains are decentralized and thus might appeal more to companies especially in the healthcare sector as they additionally also reduce the risk of information breaches by allowing only individuals who traditionally had access to a list of information to have access to it, e.g. a receptionist can only view identification information of all patients, while a caregiver can only view the medical records of their patients [52, p. 3].

4.2 Zero-Knowledge proof

As mentioned in the previous sub chapter, blockchain can improve current (mostly centralized) data handling systems, medical or otherwise, significantly by eliminating the need for an always available and fully functional central authority (e.g. a single central server) through replacing this with several decentralized nodes. Blockchain transactions are still however not completely anonymous and private as the receiver and sender blockchain addresses as well as the transaction amount are visible [53, p. 227945], publicly or in case of permissioned blockchain for all participants of the private network, which may still endanger privacy. The nodes in a network however need to have access to data in order to verify them before being added to the blockchain so making these stored data confidential by for example encrypting them makes their verification impossible by conventional methods [53, p. 227946]. This is where zero-knowledge argument schemes can be utilized to confirm the validity of a statement or the knowledge of secret values without disclosing any additional information [53, p. 227946].

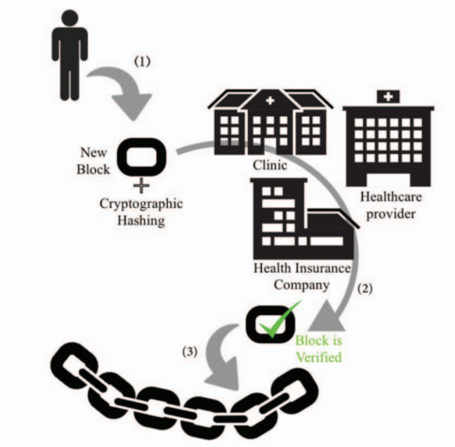


Fig. 11. Example of the working of a consortium blockchain in healthcare [52, p. 3]

The aim of zero-knowledge proof systems is for one party, the so-called prover, to convince another party, the so-called verifier, that a given statement (e.g. knowledge of some information) is valid without the prover giving the verifier any other information other than the fact that the statement is valid (or that they indeed know the information they claim) [53, p. 227947]. This needs to satisfy the following 3 properties [53, p. 227947]:

1. Completeness: A prover who has a given valid statement or knows a piece of information is able to convince the verifier of this.
2. Soundness: A malicious prover with a false statement cannot convince the verifier of this.
3. Zero-knowedge: The verifier learn nothing except the fact that the statement is true or that the prover really does know the information the prover claims he does.

In general, there are two main classes of zero-knowledge proofs: interactive Σ-protocols and Zero-knowledge succinct non-interactive arguments of knowledge (SNARKs) [53, p. 227947]. Zero-knowledge proofs are traditionally online and interactive, meaning that they require input from the verifier in the form of challenges that the prover has to solve in order to confirm their knowledge [53, p. 227947]. Assuming that there is a secure encryption scheme, all problems in the class of decision problems solvable in non-deterministic polynomial time (NP) have interactive zero-knowledge proof [53, p. 227947]. However, for the application of zero-knowledge proofs in the context of blockchain, an offline non-interactive verification is required, which only the class of decision problems solvable in bounded-error probabilistic polynomial time (BPP) possess without any additional assumptions or setups [53, p. 227947]. For all other problems, their interactive zero-knowledge proofs (if a proof exists) can be transformed into non-interactive ones through the use of different methods such as for example The common reference string (CRS) model [53, p. 227947]. The CRS model gives a non-interactive zero-knowledge proof for any problem in NP based on participants sharing a random string, and the prover computing a set result based on amongst others this common reference string as the input (instead of receiving challenges from the verifier) which the verifier then only needs to accept without ever interacting with the prover [54, p. 110-111], thus providing a non-interactive zero-knowledge proof. The aforementioned SNARKs follow this CRS model with the common reference string often generated in advance [53, p. 227947].

There are many works of literature that propose implementations of zero-knowledge proofs, most of these suggestions often combine techniques from both the aforementioned Σ-protocols and SNARKs and also borrow techniques from each other which makes a complete classification of these implementations difficult [53, p. 227948]. A possible implementation of the zero-knowledge proof for identifying blockchain users using a pair of public and private keys is shown in figure 12. Here, the user or the prover produces multiple attributes (e.g. statements) that do not compromise privacy and publishes them to a blockchain using their private blockchain key [55, p. 3]. An implemented smart contract will store the attribute and the publisher's address. Then, the identity providers use the blockchain private key to issue so-called claims (usually including the provider's signature) on the attributes of qualified users [55, p. 3]. These claims and their issuer's address are stored in the smart contract [55, p. 3]. After this preparation, the user can access the service of the service provider through demonstrating or proving that they have the identifier (controlled by their private blockchain key used to publish before), by for example making a transfer from the blockchain account wallet which allows the service provider to retrieve the user's stored identity [55, p. 3].

4.3 Laws and Regulations regarding Data Integrity and Blockchain

As mentioned in the previous chapters, it is the government of each country that is the main entity responsible for regulating social and technological issues and ensuring not only the safety of its citizens, but also that none of their basic rights (e.g. data privacy) as defined in the country's own laws or other regulations such as the International Human Rights law are violated. This also means that regulators should be able to keep up with the increasing growth of new technologies and their applications and to create and/or update the laws in this area accordingly with the best interest of the public in mind. In this chapter, we will take a closer look at the current regulations in place for Blockchain technology as well as cryptocurrencies specifically as this has been one of the main applications of the Blockchain technology in practice until recently.

Similar to section 2.3, there will also be an introduction into the regulations of the United States (as a non-EU country yet with considerable advancements

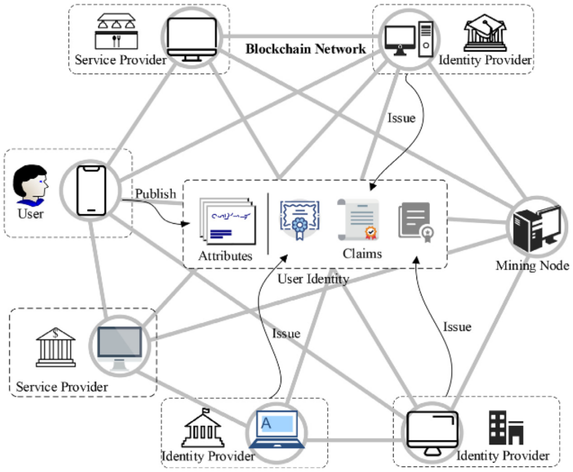


Fig. 12. Possible architecture for confirming user's identity claim [55, p. 3]

in possible usages of Blockchain in healthcare), and the European Union in the following in order to be able to compare the attitudes of countries across a range towards the Blockchain technology. There will also be a comparison of the similarities and differences in the nature as well as the extent of the regulations regarding data integrity and the blockchain technology in each country. Lastly, this chapter will also include some suggestions on how to improve the current regulation state, as this is currently not very satisfactory or extensive in some cases, through better education of lawmakers and even citizens and the involvement of experts and (educated) citizens in the process of lawmaking either directly (e.g. voting on initiatives through polls) or indirectly (e.g. through public surveys).

Data integrity, previously said to be ensured through preventing access and modification of the data by unauthorized parties, is a central concept in cybersecurity and especially important to consider with regards to the security of sensitive medical data. There are currently few regulations directly relating to or mentioning data integrity specifically. One of the main regulations relating to this concept in the US is the "Current Good Manufacturing Practice" (CGMP) Regulations by the Food and Drug Administration (FDA). The legally binding text of the CGMP is included in the title 21 of the "Code of Federal Regulations" (CFR) of the US, which is a codification of the general and permanent rules published in the federal register by governmental agencies and executives [56] whereas the rules are organized by subject. "Title 21: Food and Drugs" of the CFR contains regulations from the FDA including a description of the regulatory process by defining the requirements to be followed by drug manufacturers. The FDA defines data integrity as "the completeness, consistency and accuracy of the data" for the purpose of this regulation [57, p. 4]. This is the case when the data are “attributable, legible, contemporaneously recorded, original or a true copy, and accurate” [57, p. 4]. Attributable here means that one is able to trace back the data to the person who recorded them which prevents or at least makes clear if there has been changes by unauthorized entities. The fact that data should be attributable is ensured by a number of sections in title 21. For example, it is required in §211.101(d) that “each component shall either be added to the batch by one person and verified by a second person or, if the components are added by automated equipment under §211.68, only verified by one person” [56]. This ensures that each piece of data was written by only one person and verified at least once by a different person and makes finding out the author and verifier of a corrupt piece of data much easier in hindsight. In §211.122, it is required that all procedures regarding the packaging and labeling of material as well as any verification and approval of these procedures is documented and that this documentation is maintained properly [56]. Similarly in §211.186, it is required that the production and control records (e.g. list of components of a drug) shall be documented and these documents should be maintained by a person and verified by a second person all in written form and in §211.186(b) it should be ensured that “Documentation that each significant step in the manufacture, processing, packing, or holding of the batch was accomplished” should be included [56].

Legible here can be defined as readable, meaning that the data by one person can be understood and used later by another party, for example “written records required by this part shall be maintained so that data therein can be used for evaluating” as stated in §211.180(e) or that “All records, including those not stored at the inspected establishment, must be legible, stored to prevent deterioration or loss, and readily available for review and copying by FDA employees” as stated in §212.110(b) [56]. Contemporaneously recorded means that all documents created at the time of the performance of the drug are actually kept and recorded, and is ensured through §211.100(b) which states that “Written production and process control procedures shall be followed in the execution of the various production and process control functions and shall be documented at the time of performance” as well as through §211.160(a) which requires that “the requirements in this subpart shall be followed and shall be documented at the time of performance” [56]. The fact that the data or documents are an original or true copy is ensured in §211.180 which describes the general requirements for the written records and reports as well as in §211.194(a) which describes the requirements for laboratory records [56]. Finally, accuracy of the data is treated in §211.22(a) which requires the existence of a quality control unit and in §211.68 which defines certain control requirements for any (automatic, electronic or mechanical) equipment used throughout the process. In §211.68(b) it is specifically stated that “Appropriate controls shall be exercised over computer or related systems to assure that changes in master production and control records or other records are instituted only by authorized personnel” [56] which connects back to our previous definition of data integrity in the sense that only authorized parties should be allows to modify data. Accuracy of data is also required specifically in §211.188 for production and control records as well as in §212.60(g) where specific requirements for the laboratory performing test records are highlighted [56].

Similar to the CGMP, the European Medicines Agency (EMA) has a guideline on "good manufacturing practice" (GMP) to ensure the integrity of data that are generated in the process of testing, manufacturing, packaging, distribution and monitoring of medicines. The GMP describes the minimum requirements and standards that a manufacturer should meet in the production of medicinal products [58]. Any manufacturer of medical products intended for the EU market, no matter where the manufacturer itself is located, has to comply with the GMP. There is also a guideline on "good distribution practices" (GDP) which describes the minimum requirements that a wholesale distributor must meet to ensure the quality and integrity of medicinal products throughout the whole supply chain [59]. The GDP is legally binding for anyone engaged in wholesale distribution of medicinal products in the EEA [59]. The GMP and GDP are only scientific guidelines from the EMA, meaning that they are not legally binding. Applicants who want to enter the European market are strongly suggested to follow these guidelines but can also deviate from these if they justify their deviation in their application [60]. The GMP and GDP are both guidelines, they have however been implemented in different chapters (especially chapter 4) throughout the 4th volume of EudraLex, a set of rules governing medicinal products in the EU, and thus have become regulations throughout all member states in the EU as well as the EEA so applicants no longer can deviate freely from them. There is no specific definition of data integrity by the EMA, however the regulations in the GMP are sometimes very similar to those of the FDA. All mentioned regulations from EudraLex in the following refer to different chapters of the 4th volume of Eudralex [61]. In EudraLex §4.1 for example, it is required that "All types of document should be defined and adhered to" and that "Complex systems need to be understood, well documented, validated, and adequate controls should be in place" as well as in §4.7 which states that "Handwritten entries should be made in clear, legible, indelible way" [61] which also follows the principles of legibility as defined by the FDA. In EudraLex §4.9 for example the principle of attributability as defined by the FDa is handled as it is required that "Any alteration made to the entry on a document should be signed and dated" [61] which allows tracing the alteration back to its author in case of unauthorized or corrupt changes. This principle could even be said to have been implemented more extensively compared to the FDA regulations, as it is additionally specified in EudraLex §4.9 that "the alteration should permit the reading of the original information" and "Where appropriate, the reason for the alteration should be recorded" [61] which was not mentioned in the FDA regulations before. The accuracy principle can be seen in EudraLex regulations such as EudraLex §5.32 where an exact list of required information present on a label of starting material in the storage is given [61]. The accuracy principle from the FDA is implemented amongst others in EudraLex §6.1 which states that "Each holder of a manufacturing authorisation should have a Quality Control Department" [61] similar to §211.22(a) of title 21 of the CFR made by the FDA mentioned previously.

The definition of data integrity given at the beginning of the chapter is also confirmed in EudraLex §4.3 which states that "Documents containing instructions should be approved, signed and dated by appropriate and authorised persons" and EudraLex §4.10 requiring "Secure controls [...] to ensure the integrity of the record throughout the retention period and validated where appropriate" [61] and is also similar to the previously mentioned §211.68(b) of title 21 of the CFR. Data integrity is also mentioned in "Principle f: Integrity and confidentiality" of the GDPR which states that personal data must be "processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate technical or organisational measures" [62]. The measures which should be taken are formulated in a more broad or even vague way deliberately here as technological and organisational best practices are constantly changing [62]. This vague formulation however can also have downsides as the organizations may not feel obligated to implement the most recent and up-to-date measures or the current best practice technologies and instead opt for measures which cost less or require less work or other expenses. As it can be concluded by the above explanations, there are quite some similarities between the CGMP in the US and the GMP in the EU. Finding major differences between the two regulations however is a more difficult task especially currently as the two regulations became more compatible in the process of the "Mutual Recognition Agreement" which was reached in 2017 between the US and EU and lead to the (medical) market supplier assessments of each party being accepted by the other one [63]. This means that the successful assessment of a supplier by the US governmental organizations is now also recognized and accepted in the EU and vice versa.

There is currently no legal framework in the EU specifically concerning the processing of personal data with Blockchain or Distributed Ledger Technology as it is widely assumed that this falls under the GDPR [64, p. 13]. As we have seen previously however, the GDPR is formulated rather generally and did not specifically regulate or address Blockchain. It could even be argued that there is tension between Blockchain and the GDPR and that some variations of Blockchain may not be able to comply with the GDPR amongst others because of the way they distribute responsibility between several anonymous users or nodes and thus it would be difficult to identify the so-called "controller" as specified in the GDPR [65, p. 124]. This is however specifically the case for public blockchains and may not apply to other variations that have been developed as a response to these issues. Here it is also important to note that Blockchain technology is only a class of technology and there are several versions of this technology with sometimes widely differing functioning methods and characteristics such as public or permissioned blockchains mentioned previously. Controllership is decided based on the specific use or deployment of a technology and it could be argued that blockchain, similar to the internet, is a so-called “General Purpose Technology” that is deployed by entities for different purposes depending on the context [65, p. 125]. Therefore applying the GDPR’s notion of a controller to the internet should have posed similar issues, this is however not the case and GDPR is currently applicable to the internet so it can be concluded that the controller notion of the GDPR will also be unlikely to pose issues for blockchain [65, p. 125].

In 2019, the European Parliament published the paper “Blockchain and General Data Protection Regulation” in which some questions regarding the relation between the GDPR and Blockchain technology are discussed [66]. This study also confirms that blockchain cannot be limited or defined as one specific technology, but rather a class with variations and thus the compatibility between blockchain and the GDPR can only be determined on a case-by-case basis that accounts for the respective technical and contextual factors (such as the governance framework) [66, p. 7]. The study does discusses some areas where there might be tension between some blockchain variations and the GDPR based on examples, however it also highlights how the technology could help better achieve some of the GDPR’s underlying objectives [66, p. 91]. Recital 7 of the GDPR for example states that “natural persons should have control of their own personal data” which is implemented in Article 15 of the GDPR (right of access) and in Article 20 of the GDPR (the right to data portability) amongst others [66, p. 92]. Having control in this context can thus be defined as each person being able to monitor what happens to personal data relating to them and to decide who can have access to this data which can be difficult to achieve in practice [66, p. 92]. Blockchain could however be of use to increase the control of the data subjects over their own personal data due to characteristics such as transparency (regarding who has accessed data) and decentralized data-sharing possibility (without the need for a trusted central authority). A technical infrastructure similar to blockchain has for example already long been used in Estonia to provide data subjects with more control over their health data by enabling them to assess all authorizations regarding their data [66, p. 92]. Medical specialists can by default access a patient’s data but the patient can even choose to deny a certain specialist or all specialists access to a certain case data [66, p. 92]. The application of blockchain as a control-bestowing tool especially regarding health data is currently being explored further and some projects have been developed to for example research the potential of data sharing solutions based on blockchain in the health sector [66, p. 92-93]. For example Patientory is a smartphone app which uses distributed ledgers to encrypt and shred electronic health records to prevent data breaches and MedRec is a mobile app which uses smart contracts as a record management system for electronic medical records in multi-institutional settings [66, p. 93]. This is an optimistic prospect for potential future application of the Blockchain technology in other areas and especially in healthcare which may improve the status quo not only from a technical, but even from a regulatory aspect by granting users more control amongst others and generally helping to achieve the objectives of regulations such as the GDPR better and more efficiently. The "Blockchain and General Data Protection Regulation" paper is similarly a first step to legal certainty regarding blockchain, however it cannot be considered as a sufficiently reliable legal framework and more exact legislations and care studies should follow soon [64, p. 13].

Similarly, the research to find any legal framework by official governmental institutions in the US yielded no regulations or guidelines in regard to blockchain technology specifically without the mention of cryptocurrencies and blockchain-based payment systems. It is also interesting that similar issues regarding the compatibility of blockchain and the GDPR have been raised also for blockchain and HIPAA [67]. It seems that both in the EU and the US there is currently some blockchain-specific regulations only in the context of cryptocurrencies as this is one of the first and oldest applications of the technology and is still currently in use today. Trades and exchanges involving virtual currencies are subject to the Commodity Exchange Act (CEA) issued by the Commodity Futures Trading Commission (CFTC) in the US [68]. There was also a "Digital Commodity Exchange Act" issued in 2020 specifically to regulate the exchange of digital commodities such as cryptocurrencies [69]. A new bill to improve this act has been proposed recently in April 2022 by the Congress with the aim to give the CFTC a bigger role in overseeing crypto spot markets [70]. The regulation of cryptocurrencies in the US is still relatively divided across a range of different laws and (federal) agencies depending on the use-case (e.g. selling, buying or exchanging) of crypto assets. The sale of cryptocurrencies is generally only regulated if it concerns the sale of a security or if the sale is considered money transmission making the seller a so-called "Money Services Business" (MSB) [68]. The Securities Exchange Commission (SEC) is in charge of regulating the issuing or reselling any digital assets that fall under the definition of a security [68]. Under the US law, an investment contract, namely a monetary investment in a common enterprise with a reasonable expectation of profits to be derived from entrepreneurial or managerial efforts of others, also counts as a security [68]. In case of a digital asset, whether or not it is an "investment contract" is decided based on not the form but the substance of the transaction [68]. The Financial Crimes Enforcement Network (FinCEN) regulates cryptocurrencies which may be considered MSBs under the Bank Secrecy Act (BSA) [68]. According to a guidance issued by FinCEN in 2013, both a "virtual currency exchanger" as well as "an administrator of a centralized repository of virtual currency who has the authority to both issue and redeem the virtual currency" fall under the definition of an MSB [68]. An MSB that is a money transmitter must conduct a risk assessment of its exposure to money laundering and implement an anti-money laundering (AML) program. based on this assessment [68]. Under FinCEN's regulations, an administration or exchanger that "accepts and transmits a a convertible virtual currency" or "buys or sells convertible currency for ny reason" is a money transmitter [68]. The taxation of cryptocurrencies falls under the jurisdiction of the Internal Revenue Service (IRS) and cryptocurrencies are considered as “property” for taxation purposes according to Section 4 of the IRS Notice 2014-21 titled “IRS Virtual Currency Guidance” published in 2014 [71]. Therefore, every individual or business that own cryptocurrency needs to keep detailed records of their purchases and sales with cryptocurrencies and pay taxes on any gain that may have been made upon the sale of cryptocurrency for cash or made upon the purchase of a good or service with cryptocurrency and pay taxes on the fair market value of the cryptocurrency on the date of receipt [68].

Within the EU, the regulation of cryptocurrencies is also split across a number of governmental organizations and different legislations. Crypto-assets which qualify as “e-money” or “financial instruments” are regulated by the exiting EU financial services legal framework. “E-money” is defined in Article 2.2 of the Electronic Money Directive (EMD) as any “electronically, including magnetically, stored monetary value as represented by a claim on the issuer which is issued on receipt of funds for the purpose of making payment transactions [...] and which is accepted by a natural or legal person other than the electronic money issuer” [72]. A list of activities qualifying as a “financial instrument” is given in Article 4.1.3.15 of the Markets in Financial Instruments and amending Directive (MIFID) [73]. Legislation applying to the issuance, trading, clearing and settlement of financial instruments is largely harmonized at EU level through regulations such as the aforementioned MIFID or the Central Security Depository Regulation [74, p. 2] which regulates the transactions of securities as defined previously. This leaves little room for the EU member states to develop their own legal framework on a national level.

According to a study supported by the European Parliament, one of the key issues in regulating and monitoring transactions involving cryptocurrencies is the anonymity of the involved parties which facilitates unethical transactions as well as tax evasion for criminal organizations amongst others [75, p. 9]. The European Union has long aimed to fight money laundering and terrorism financing, adopting its first Anti-Money Laundering Directive (AMLD1) on the 10th of June 1991 in order to coordinate measures across different member states and ensure the stability of the financial framework within the EU as a whole [75, p. 58]. The AMLD has since been reviewed and updated four times with the fifth Anti-Money Laundering Directive (AMLD5), which came into effect in January 2020, expanding the scope of AMLD4 to also include cryptocurrencies. Article 2.1.3 of the AMLD4 listing the “obliged entities” which the directive applies to has been updated by AMLD5 to additionally include “providers engaged in exchange services between virtual currencies and fiat currencies” and “custodian wallet providers” [76]. “Custodian wallet provider” is later defined in Article 3.19 of AMLD5 as “entity that provides services to safeguard private cryptographic keys on behalf of its customers, to hold, store and transfer virtual currencies” [76]. AMLD5 also defines virtual currencies in Article 3.18 as “a digital representation of value that is not issued or guaranteed by a central bank or a public authority, is not necessarily attached to a legally established currency and does not possess a legal status of currency or money, but is accepted by natural or legal persons as a means of exchange and which can be transferred, stored and traded electronically” [76]. Thus, AMLD5 subjects also virtual currency exchange services and custodian wallet providers to legal evaluations and requires them to report suspicious transactions which leads to easier identification and recognition of money laundering activities or tax evasion involving virtual currencies. However, it is important to note that a number of key players in cryptocurrency markets such as miners, hardware or software wallet providers or trading platforms and coin offerors are still not included in the AMLD5 which leaves blind spots that can misused by malicious parties [75, p. 9].

Generally, it can be noted that the regulation of Distributed Ledger Technology or Blockchain still has a long way to go. Even regulations for its application in context of cryptocurrencies, which is a technology that has been around for over a decade now (starting with Bitcoin in 2008 [47]), are still lacking and do not cover all stakeholders and players in the cryptocurrency market. A lack of regulations regarding Blockchain in general and outside of cryptocurrencies can also be observed in the EU and the US based on the above research. This may to some extent be due to the difficulties in finding a unified definition for the term Blockchain. This term is very broad as it includes several different variations of the technology and keeps changing and evolving to include new variants relatively quickly [66, p. 7] so a more permanent definition suitable for legal documents is difficult to find. Lawmakers should nonetheless attempt to regulate at least the existing applications of the Blockchain technology and prepare to also regulate its future applications which are emerging. The use of Blockchain in healthcare is for example increasing and it should be considered for future regulations to ensure the safety and protection of all stakeholder including the patients and to prevent or reduce considerable damage and loss due to misuse or malicious attacks. In order to better regulate technologies, it is important that governments and lawmaking authorities familiarize themselves with these technologies and their most important underlying concepts in addition to getting aid and consultation from appropriate experts. Understanding the technologies and their applications will allow lawmakers to not only create suitable regulations, but to also identify blind spots in existing regulations and improve these. It is also important to educate and involve citizens and to get their feedback (or that of qualified representatives of groups of citizens) on regulations (e.g. through voting polls or consensus conferences). It has for example been noted that with cryptocurrencies, the consumers’ lack of understanding of the underlying technology may significantly increase operational risks and the risk of fraud [74, p. 2]. Alternatively, the concept of Scientific Lawmaking could also be applied here which divides involved parties in the lawmaking process into legislatures, which issue policies, and institutions, which design laws based on these policies [77]. Legislators here will not be involved in the design of of laws, but instead will identify problems, discuss issues and set policies but will assign the design of laws based on their policies to qualified law-design institutions [77].

5 Conclusion

5.1 Summary and Conclusion

In this work, the current state, technologies and some of the trends in healthcare, as well as some of the most used cybersecurity measures were discussed and some of the weaknesses of these technologies and measures in the context of their application in healthcare were highlighted (e.g. relatively easy manipulation through social engineering techniques such as phishing). The focus here was to provide an overview of the most common technologies and security measures currently in use as well as trends which have been researched and discussed quite often in literature. This work also contributes to a clearer overview of the main data privacy regulations currently being enforced in the US, EU and EEA such as the the HITECH or GDPR. Additionally, two central concepts, resilience and integrity as well as methods to implement these concepts were introduced. Resilience in healthcare has only gained attention in the recent years and some facilities still lack initiatives and measures to implement this concept into their operation. This work introduces some of the suitable available models for more resilient healthcare services (especially in everyday services but also during shocks such as pandemics or natural disasters), however there are also other models for implementing resilience being developed every day by experts and researchers which were not mentioned in this work as introducing all concrete models was not the goal of this work. Data integrity is one of the main focus points, where this work contributes to approaches to implementing data integrity by highlighting the potential of the use of the Distributed Ledger Technology (DLT) or Blockchain. Due to the (immutable) nature of this technology, some of its variations such as permissioned Blockchain are especially suitable for ensuring healthcare data integrity without much (or any) additional work or expenses. Finally, an overview of the current regulations regarding the Blockchain technology and its current applications (e.g. cryptocurrencies) was given where it was concluded that there are not sufficient regulations for Blockchain generally (perhaps due to difficulties in defining and identifying this dynamic concept and its changing variations). Existing regulations mostly exist for the application of Blockchain in the context of cryptocurrencies (possibly because this was one of the first applications of Blockchain) but even these regulations should be improved in some aspects as this technology and its use transform over time.

5.2 Limitations and Scope of the Work

The focus of this work was on data resilience and integrity, arguably the two most important aspects that should be ensured in the protection of sensitive data. But ensuring the quality and security of data is a multi-dimensional concept and is difficult to narrow down to a limited number of elements so it should be acknowledged that there are also other aspects besides integrity and resilience which may be looked into. It should additionally be noted that there are also other technologies used in healthcare currently and as mentioned above many more will be developed and used in the future with advancements in technology. This work only discusses some of the most commonly used concepts and technologies and the trends discussed should be regarded in the context of the time this work was written.

Additionally, while the most important data privacy regulations were introduced, the aim of this work was not to provide an extensive list of all Articles concerning data privacy within every regulation, but to provide an overview of the most important regulations within each of the chosen countries or groups of countries and to compare and contrast the regulatory and legal situation of these (groups of) countries with each other. The United States of America (referred to as the US throughout the work) was chosen as it has one of the more discussed and well-known healthcare systems and as a non-European country for better comparison and to provide a different perspective. However, all points discussed and the comparison naturally only applies to the US and not all non-European countries as each country has its own individual healthcare system and privacy regulations. The European Union (EU) was chosen as its regulations cover a large group of countries (all member countries) and even affect some other countries with relation to the EU (e.g. Norway as an EEA) country. It was also of interest to discuss some differences between the EU and the EEA regulations which are often assumed to be identical or very similar. Similar to the comparison with the US, the points discussed about the technological state of the healthcare systems in Norway and Austria may not apply to all other EU and EEA countries and the comparisons cannot be generalized.

5.3 Future Work

This paper has given an overview of the strengths and weaknesses of the current health information systems working with electronic medical records and the data privacy regulations concerning healthcare. While some methods and possible measures to improve these deficiencies were suggested throughout the work, concrete models and step-by-step measures to improve specific weaknesses of the current technologies and regulations may be developed in the future.

The potential of the Blockchain technology for ensuring data integrity, also in the healthcare sector, was explored in this paper. Similarly, there is the possibility of developing specific models for integrating Blockchain into different types of IT-Systems currently in use in healthcare. While there are some models developed for integrating Blockchain into a system of existing security measures (e.g. Access Control), there is still much to be researched especially regarding the the use of Blockchain in specific medical facilities and IT-systems.

List of Figures

References

1. The National Coordinator for Health Information Technology (ONC), "Benefits of ehrs." https://www.healthit.gov/topic/health-it-and-health-information-exchange-basics/benefits-ehrs. Accessed: 2022-07-20.
2. Critical Insight, "Healthcare breach report july-dec 2021." https://cybersecurity.criticalinsight.com/2021_H2_HealthcareDataBreachReport.
3. N. P. Walsh, "Serious cyberattacks in europe doubled in the past year," Jun 2021.
4. United Nations (UN), "The universal declaration of human rights (udhr)." https://www.un.org/en/about-us/universal-declaration-of-human-rights. Accessed: 2022-07-21.
5. The European Parliament, "Charter of fundamental rights of the european union of 26 october 2012." `https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT`.
6. United Nations Development Programme (UNDP), "Health information systems." https://www.undp-capacitydevelopment-health.org/en/capacities/focus/health-information-systems/. Accessed: 2022-07-22.
7. M. Müllhauser, "Introduction to Ubiquitous Computing," in *Handbook of Research on Ubiquitous Computing Technology for Real Time Enterprises* (M. Müllhauser and I. Gurevych, eds.), pp. 1–20, IRS, 2008.
8. S. B. Baker, W. Xiang, and I. Atkinson, "Internet of things for smart healthcare: Technologies, challenges, and opportunities," *IEEE Access*, vol. 5, pp. 26521–26544, 2017.
9. S. M. R. Islam, D. Kwak, M. H. Kabir, M. Hossain, and K.-S. Kwak, "The internet of things for health care: A comprehensive survey," *IEEE Access*, vol. 3, pp. 678–708, 2015.
10. A. Tahir, F. Chen, H. U. Khan, Z. Ming, A. Ahmad, S. Nazir, and M. Shafiq, "A systematic review on cloud storage mechanisms concerning e-healthcare systems," *Sensors*, vol. 20:5392, Sep 2020.
11. A. Rai, "What is big data – characteristics, types, benefits & examples," May 2020. `https://www.upgrad.com/blog/what-is-big-data-types-characteristics-benefits-and-examples/#Characteristics_of_Big_Data`. Accessed: 16. April 2022.
12. R. Agrawal and S. Prabakaran, "Big data in digital healthcare: Lessons learnt and recommendations for general practice," *Heredity*, vol. 124, p. 525–534, Mar 2020.
13. A. Arora, "Conceptualising artificial intelligence as a digital healthcare innovation: An introductory review," *Medical Devices*, vol. Volume 13, p. 223–230, Aug 2020.
14. H. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. Hassen, L. Thomas, A. Enk, and et al., "Man against machine: Diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, p. 1836–1842, May 2018.
15. A. J. Steele, S. C. Denaxas, A. D. Shah, H. Hemingway, and N. M. Luscombe, "Machine learning models in electronic health records can outperform conventional survival models for predicting patient mortality in coronary artery disease," *PLOS ONE*, vol. 13, Aug 2018.
16. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Y. Ding, A. Bagul, C. P. Langlotz, K. S. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, vol. abs/1711.05225, 2017.

17. S. S. Bhuyan, U. Y. Kabir, J. M. Escareno, K. Ector, S. Palakodeti, D. Wyant, S. Kumar, M. Levy, S. Kedia, D. Dasgupta, and A. Dobalian, "Transforming healthcare cybersecurity from reactive to proactive: Current status and future recommendations," *Journal of Medical Systems*, vol. 44:98, 2020.
18. D. Rountree, "4 - system security," in *Security for Microsoft Windows System Administrators*, pp. 109–134, Boston: Syngress, 2011.
19. U. D. of Health and H. Services, "Health sector cybersecurity: 2021 retrospective and 2022 look ahead." March 3, 2022. `https://www.hhs.gov/sites/default/files/2021-retrospective-and-2022-look-ahead-tlpwhite.pdf`.
20. Radware, "Anyone is a target: Dos attack case analysis on boston children's hospital." `https://www.radware.com/getattachment/Security/ERT-Case-Studies/771/Radware_Boston_Childrens_Hospital_Case_Study.pdf.aspx/?lang=en-US`. Accessed: 2022-05-22.
21. K. Abouelmehdi, A. Beni-Hssane, H. Khaloufi, and M. Saadi, "Big data security and privacy in healthcare: A review," *Procedia Computer Science*, vol. 113, pp. 73–80, 2017. The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017) / The 7th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH-2017) / Affiliated Workshops.
22. S. Shah and S. Kanhere, "Recent trends in user authentication - a survey," *IEEE Access*, vol. PP, pp. 1–1, 08 2019.
23. A. Ghazvini and Z. Shukur, "Review of information security guidelines for awareness training program in healthcare industry," in *2017 6th International Conference on Electrical Engineering and Informatics (ICEEI)*, pp. 1–6, 2017.
24. C. S. Kruse, B. Frederick, T. Jacobson, and D. K. Monticone, "Cybersecurity in healthcare: A systematic review of modern threats and trends," *Technology and health care : official journal of the European Society for Engineering and Medicine*, vol. 25, pp. 1–10, 08 2016.
25. C. S. Kruse, B. Smith, H. Vanderlinden, and A. Nealand, "Security techniques for the electronic health records," *Journal of Medical Systems*, vol. 41: 127, 07 2017.
26. HIPAA Journal, "https://www.hipaajournal.com/what-is-the-hitech-act/." Accessed: 2022-05-24.
27. Public Law 111–148 (The Patient Protection and Affordable Care Act) of the 111th Congress enacted by the Senate and House of Representatives of the USA, "https://www.healthcare.gov/where-can-i-read-the-affordable-care-act/." Accessed: 2022-05-26.
28. U.S. Food and Drug Administration, "https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device." Accessed: 2022-05-25.
29. M. Adibuzzaman, P. DeLaurentis, J. Hill, and B. Benneyworth, "Big data in healthcare - the promises, challenges and opportunities from a research perspective: A case study with a model database," *AMIA (Annual Symposium proceedings)*, p. 384–392, 2017. PMID: 29854102; PMCID: PMC5977694.
30. Nate Lord, "What is the data protection directive? the predecessor to the gdpr." https://digitalguardian.com/blog/what-data-protection-directive-predecessor-gdpr, 2018-09-12. Accessed: 2022-05-26.
31. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, "http://data.europa.eu/eli/dir/1995/46/oj." Accessed: 2022-05-26.

32. Regulation (EU) 2016/679 of The European Parliament and The Council of The European Union of 27 April 2016, “https://gdpr-info.eu/.” Accessed: 2022-05-26.
33. Volker Schörghofer, “Fünf jahre e-card: Entwicklungen und auswirkungen auf das verhalten der user.” https://www.chipkarte.at/cdscontent/load?contentid=10008.551367&version=1391172970, 2011-05. Accessed: 2022-05-28.
34. oesterreich.gv, “System der pflichtversicherung (asvg, gsvg, fsvg, bsvg).” https://www.oesterreich.gv.at/themen/arbeit_und_pension/pension/1/Seite.270110.html. Accessed: 2022-05-28.
35. Bundesgesetz vom 9. September 1955 über die Allgemeine Sozialversicherung (Allgemeines Sozialversicherungsgesetz – ASVG.), “https://www.ris.bka.gv.at/geltendefassung.wxe?abfrage=bundesnormen&gesetzesnummer=10008147.” Accessed: 2022-05-28.
36. Rune Opdahl and Pernile Gjerde Lia, “Norway - data protection overview.” https://www.dataguidance.com/notes/norway-data-protection-overview. Accessed: 2022-06-01.
37. Norsk Helsenett, “https://www.helsenorge.no/.” Accessed: 2022-06-01.
38. Direktoratet for E-Helse, “https://www.ehelse.no/prosjekt/helseplattformen.” Accessed: 2022-06-01.
39. LOV-2021-06-18-124: Lov om behandling av personopplysninger (personopplysningsloven), “https://lovdata.no/dokument/nl/lov/2018-06-15-38.” Accessed: 2022-06-01.
40. LOV-2000-04-14-31: Lov om behandling av personopplysninger (personopplysningsloven), “https://lovdata.no/dokument/nlo/lov/2000-04-14-31.” Accessed: 2022-06-01.
41. D. Essuman, N. Boso, and J. Annan, “Operational resilience, disruption, and efficiency: Conceptual and empirical analyses,” *International journal of production economics*, vol. 229:107762, Apr 2022.
42. R. Caralli, J. Allen, P. Curtis, D. White, and L. Young, “Cert resilience management model, version 1.0,” Tech. Rep. CMU/SEI-2010-TR-012, Software Engineering Institute, Carnegie Mellon University, Pittsburgh, PA, 2010.
43. H. B. Lyng, C. Macrae, V. Guise, C. Haraldseid-Driftland, B. Fagerdal, L. Schibevaag, and S. Wiig, “Capacities for resilience in healthcare; a qualitative study across different healthcare contexts,” *BMC Health Services Research*, vol. 22, Apr 2022.
44. J. E. Anderson, A. J. Ross, J. Back, M. Duncan, P. Snell, K. Walsh, and P. Jaye, “Implementing resilience engineering for healthcare quality improvement using the care model: A feasibility study protocol,” *Pilot and Feasibility Studies*, vol. 2, Oct 2016.
45. M. Zviran and C. Glezer, “Towards generating a data integrity standard,” *Data and Knowledge Engineering*, vol. 32, pp. 291–313, 03 2000.
46. A. K. Pandey, A. I. Khan, Y. B. Abushark, M. M. Alam, A. Agrawal, R. Kumar, and R. A. Khan, “Key issues in healthcare data integrity: Analysis and recommendations,” *IEEE Access*, vol. 8, pp. 40612–40628, 2020.
47. Satoshi Nakamoto, “Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf.” 2008, Accessed: 2022-06-03.
48. R. Schollmeier, “A definition of peer-to-peer networking for the classification of peer-to-peer architectures and applications,” in *Proceedings First International Conference on Peer-to-Peer Computing*, pp. 101–102, 2001.

49. L. M. Bach, B. Mihaljevic, and M. Zagar, "Comparative analysis of blockchain consensus algorithms," in *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pp. 1545–1550, 2018.
50. D. D. F. Maesa, P. Mori, and L. Ricci, "A blockchain based approach for the definition of auditable access control systems," *Computers & Security*, vol. 84, pp. 93–119, 2019.
51. B. Liu, X. L. Yu, S. Chen, X. Xu, and L. Zhu, "Blockchain based data integrity service framework for iot data," in *2017 IEEE International Conference on Web Services (ICWS)*, pp. 468–475, 2017.
52. Z. Alhadhrami, S. Alghfeli, M. Alghfeli, J. A. Abedlla, and K. Shuaib, "Introducing blockchains for healthcare," in *2017 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*, pp. 1–4, 2017.
53. J. Partala, T. H. Nguyen, and S. Pirttikangas, "Non-interactive zero-knowledge for blockchain: A survey," *IEEE Access*, vol. 8, pp. 227945–227961, 2020.
54. M. Blum, P. Feldman, and S. Micali, "Non-interactive zero-knowledge and its applications," STOC '88, (New York, NY, USA), p. 103–112, Association for Computing Machinery, 1988.
55. X. Yang and W. Li, "A zero-knowledge-proof-based digital identity management scheme in blockchain," *Computers & Security*, vol. 99, p. 102050, 2020.
56. The Electronic Code of Federal Regulations (eCFR), "https://www.ecfr.gov/."
57. Food And Drug Administration (FDA), "Data integrity and compliance with drug cgmp." https://www.fda.gov/media/119267/download. December 2018.
58. European Medicines Agency (EMA), "Good manufacturing practice." https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-manufacturing-practice.
59. European Medicines Agency (EMA), "Good distribution practice." https://www.ema.europa.eu/en/human-regulatory/post-authorisation/compliance/good-distribution-practice.
60. European Medicines Agency (EMA), "Scientific guidelines." https://www.ema.europa.eu/en/human-regulatory/post-authorisation/compliance/good-distribution-practicehttps://www.ema.europa.eu/en/human-regulatory/research-development/scientific-guidelines.
61. European Commission - Public Health, "Eudralex - volume 4." `https://health.ec.europa.eu/medicinal-products/eudralex/eudralex-volume-4_en`. 2011.
62. L. Irwin, "The gdpr: Understanding the 6 data protection principles." https://www.itgovernance.eu/blog/en/the-gdpr-understanding-the-6-data-protection-principles. December 2021. Accessed: 2022-06-20.
63. European Medicines Agency (EMA), "Mutual recognition agreements (mra)." https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-manufacturing-practice/mutual-recognition-agreements-mra.
64. European Crowdfunding Network (ECN), "Analysis of current european blockchain regulation." `https://eurocrowd.org/wp-content/uploads/2021/03/20201006_Analysis_of_current_European_Blockchain_Regulation_ECN.pdf`.
65. L. Moerel and M. Storm, "Blockchain can both enhance and undermine compliance but is not inherently at odds with eu privacy laws," *Journal of Investment Compliance*, vol. 22, p. 122–132, Apr 2021.
66. M. Finck, "Blockchain and the general data protection regulation pe 634.445," *European Parliamentary Research Service*, July 2019.

67. M. Miliard, “Blockchain faces tough roadblocks in healthcare,” April 2017. Healthcare IT News. https://www.healthcareitnews.com/news/blockchain-faces-tough-roadblocks-healthcare.
68. J. Dewey, “Blockchain and cryptocurrency laws and regulations 2022: Usa,” October 2021. Global Legal Insights. https://www.globallegalinsights.com/practice-areas/blockchain-laws-and-regulations/usa.
69. 116th Congress of the US, “H.r.8373 - digital commodity exchange act of 2020.” https://www.congress.gov/bill/116th-congress/house-bill/8373/text?r=5&s=1.
70. E. Harp, “Congress introduces digital commodity exchange act,” April 2022. ETF Database. https://etfdb.com/crypto-channel/congress-introduces-digital-commodity-exchange-act/. Accessed: 2022-07-02.
71. Internal Revenue Service (IRS), “Internal revenue bulletin: 2014-16,” April 2014. https://www.irs.gov/irb/2014-16_IRB. Accessed: 2022-07-02.
72. The European Parliament, “Directive 2009/110/ec of 16 september 2009 on the taking up, pursuit and prudential supervision of the business of electronic money institutions.” https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32009L0110.
73. The European Parliament. Directive 2014/65/EU of 15 May 2014 on markets in financial instruments. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014L0065.
74. The European Commission. Inception Impact Assessment on Directive/regulation establishing a European framework for markets in crypto assets. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12089-Financial-services-EU-regulatory-framework-for-crypto-assets_en.
75. R. Houben and A. Snyers, “Cryptocurrencies and blockchain: Legal context and implications for financial crime, money laundering and tax evasion pe 619.024,” *European Parliamentary Research Service*, July 2018.
76. The European Parliament. Directive (EU) 2018/843 of the European Parliament and of the council of 30 May 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32018L0843.
77. The Science of Laws Institute, “Scientific lawmaking.” https://scienceoflaws.org/the-solution/scientific-lawmaking.aspx. Accessed: 2022-07-13.