Karlsruhe Institute of Technology

Where did my data go? Evaluation of Distributed Ledger Technologies' Suitability for Personal Data Provenance

Bachelor's Thesis of

Aleksandar Bachvarov

at the Department of Informatics, Institute of Information Security and Dependability (KASTEL)
Decentralized Systems and Network Services Research Group

Reviewer:	Prof. Dr. Hannes Hartenstein
Second reviewer:	Prof. Dr. Ali Sunyaev
Advisor:	M.Sc. Oliver Stengele
Second advisor:	M.Sc. Jan Bartsch

01. Oct 2021 – 01. Feb 2021

I declare that I have developed and written the enclosed thesis completely by myself, and have not used sources or means without declaration in the text.

PLACE, DATE

. . . . . .

(Aleksandar Bachvarov)

Contents

1 Introduction

1.1 Current State

With e-health [Eys01], e-finance [AMS02], cloud services, 'Internet of Things', social media, etc. spreading and growing by the day, data exchanged, analysed or produced by intelligent devices become more and more difficult to trace [17]. It is often unknown how information is collected, how it is further processed, by whom, and for what purpose [Zub15]. This kind of information is often referred to as *data provenance* (DP), where "The provenance of a data item includes information about the processes and sources that lead to its creation and current representation" [GD07, p. 3]. The purpose of provenance is to extract relatively simple explanations for the existence of some piece of data from some complex workflow of data manipulation.

With digitalisation, the concern with potential exposure of private and sensitive personal information is rising [TQV21], and with it, the significance of DP [BT19]. Also, information is not only personal and private, but also proprietary. Consumers should know if their data had been manipulated and how, in a network, that provides interoperability and connects actors in a secure, trustworthy and 'user friendly' way [Sun+14].

An increasing amount of research is being done to utilize DP technologies [BT19] in the fields of *healthcare* [Mar+20; LAC19; Le 18; HK21; Rah+20; Sun+14], *finance* [Sin+20; Liu+21; SAD19; Sir+19], supply-chain [Man+18], cloud services [Xia+17], scientific research [SPG05], storage systems [Mun+06], etc.

A lot of progress has been made recently regarding personal data and its protection [; 18; 19, TRND]. In European data protection law, everybody has the right to know where the organisation accountable got his data from, what the data was used for, where it was transferred to and how long it is stored, regardless of location [, GDPR]. However, laws and regulations alone cannot provide consumers with information about their personal data [CAG02]. The regulations created the need for tools, which can enable consumers to exercise their rights.

Unfortunately, many tools failed to meet the requirements of such technology [Hed08; Nor09; Hu+20]. In order for such tools to work, a combination of not only proper standards and legislation is needed, but also international adoption as well as mature and suitable technologies and architectures for their development [CAG02]. When improperly designed, DP tools can be a severe threat to the consumer and in a networked environment with a lot of actors this can be a complex and costly system to implement and manage [Hed08].

There are tools that partially solve some of the existing problems like owning your data, knowing where it is stored and what's happening to it [, MTM], others provide full access to all personal data along information flows [BKB16] or easy-to-understand visualization techniques [SS17]. However, these tools are still built in a centralised manner. While centralised databases provide advantages in terms of, for instance, maintainability, they have drawbacks in terms

of their availability, performance (bottlenecks), and don't necessarily solve the issue with untrustworthiness [Sun20, p. 266-267].

1.2 Proposed Solution

To desire a one-fits-all solution is unrealistic. Recently, however, the *distributed ledger technologies* (DLTs) are on the rise and steadily becoming more versatile in terms of applicable use cases [Mau+17]. DLT has been developed to keep a distributed immutable ledger of financial transactions [Sun20]. The ledger can be seen as a provenance record of, say, bitcoins; and it is therefore unsurprising that DLT could be used to record provenance in other settings. There are many fields, which process data of sensitive and personal nature. However, in this work we will focus on the domain of *healthcare* and *finance*, as examples of domains that, although both dealing with private or personal information, still have different goals, scope, significance, etc. For instance, *healthcare* is in the public sector, while *finance* is in the private; *healthcare* is largely dominated by non-profit organisations, while *finance* has investor-owned businesses; *healthcare* payments are made by insurance companies or the government, while in *finance* usually the consumer is managing his own expenses.

However, one thing in common in *healthcare* and *finance*, for example, is consumers' and patients' trust, that their personal information is protected and safe. This can be, therefore, seen as one of the most important requirements for a personal DP approach.

By leveraging the global-scale computing power of distributed networks, a DLT-based DP can provide integrity, authenticity, traceability, accountability, provenance, trustworthiness and more through its decentralized architecture, immutable record of transactions, lack of single authority, consensus mechanisms, smart contracts, tamper-proof storage of data, etc. [; Mar+20; Mun+06] and, thus, solve the issue with untrustworthiness and fulfil important requirements of DP approaches.

There are, however, different DLTs and they vary from each other in many ways such as their design, purpose, way of access, way of governance and so on [Cho+19]. So it is important to understand the characteristics, capabilities and trade-offs of individual DLTs [Kan+20] in order to determine the most suitable approaches for personal DP in the field of *healthcare* and *finance*. This leads us to the research question:

RQ: What are the characteristics of DLTs that make them suitable for personal data provenance in healthcare and finance?

1.3 Outline

...

2 Fundamentals

2.1 Data Provenance

Data Provenance (DP) - In this work we define DP as an approach/technology that can be used to record *personal data*. This definition includes not only metadata, data origin and/or data operation, but also processes that act on data and agents that are responsible for those processes (sharing, storage, exhange, access). Most importantly, this should be achieved in a secure, trustworthy and traceable way, that ensures accountability and is in accordance to international laws and regulation, with the well-being of the consumer in mind.

Personal Data - 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 [Tru18].

2.2 Distributed Ledger Technology

A distributed ledger (also called a shared ledger or distributed ledger technology or DLT) is a consensus of replicated, shared, and synchronized digital data geographically spread across multiple sites, countries, or institutions [Sun20]. Unlike with a centralized database, there is no central administrator [Sca16].

The distributed ledger database is spread across several devices (nodes) on a peer-to-peer network, where each replicates and saves an identical copy of the ledger and updates itself independently. The primary advantage is the lack of central authority. When a ledger update happens, each node constructs a new transaction, and then the nodes vote by consensus algorithm on which copy is correct. Once a consensus has been determined, all the other nodes update themselves with the new, correct copy of the ledger [Mau+17]. Security is accomplished through cryptographic keys and signatures [Sun20]. The literature [Kan+20] differentiates between:

DLT *concepts* - describe the basic structure and functioning of DLT designs on a high level of abstraction. For instance, blockchain is a DLT concept describing the use of blocks that form a linked list. Each block contains multiple transactions that have been added into the block by nodes [Kan+20].

DLT *designs* - specify an abstract description of DLT concepts by adding concrete values and processes for inherent DLT characteristics. There are important differences between DLT

designs, which make them suitable for some applications and unsuitable for others [Kan+20].

DLT *characteristics* - represent features of DLT designs, which are of technical or administrative nature. The technical characteristics constrain future changes of the administrative characteristics(e.g., lack of scalability regarding network size of a distributed ledger) [Kan+20].

DLT *properties* - groups of DLT characteristics and shared by each DLT design. For instance, "throughput" and "scalability" are both associated with the DLT property "performance" [Kan+20].

The emergence of DLT, with strong support for data integrity, authenticity and provenance, has opened up the door of opportunities in different domains [; Mar+20; Mun+06; Lia+17; Wor+20]. With the increase in DLT application domains, the number of DLT designs has also increased steadily. These DLT designs vary from each other in many ways such as implementation, purpose, way of access, way of governance and so on [Cho+19]. Therefore, it is important to understand the characteristics of DLT designs and their properties, in order to determine which are more advantageous and most importantly, which properties make them suitable (or not) for a particular use case and its specific requirements.

2.2.1 Designs

DLT designs can be instantiated as a *public* or *private*, which can be further divided into *permissioned* and *permissionless* [Xu+17; Yeo+17].

Public - DLT designs, where the underlying network allows arbitrary nodes to join and participate in the distributed ledger's maintenance. For example, consumers can execute financial transaction without registration or verification of the nodes' identities being required. Public DLT designs like, for example, Ethereum [] are usually maintained by a large number of nodes. Owing to the large number of nodes in the network, each of which stores a replication of the ledger, public DLT designs achieve a high level of availability. To allow many (arbitrary) nodes to find consensus, public DLT designs should be well scalable to not deter performance when the number of nodes increases [Sun20].

Private - DLT designs, that engage a defined set of nodes, with each node identifiable and known to the other network nodes. Consequently, private DLT designs require verification of the nodes that join the distributed ledger. Private DLT designs like, for example, Hyperledger [DMH17] are often used if the public should not be able to access the stored data [BM16]. For example, physicians can use a common ledger in Healthcare to collaborate, but do not want to disclose the data to other colleagues or institutions not involved in the collaboration [Sun20].

Permissioned - when consensus finding is delegated to a subset of nodes (which is usually small). Since only selected nodes can validate new transactions or participate in consensus finding, fast consensus finding can be applied, which enables a throughput of multiple thousands of transactions per second [CL+99]. Owing to the small number of nodes involved in consensus finding, they can reach finality, which means that all of a distributed ledger's permitted nodes

come to an agreement regarding the distributed ledger's current state [Sun20].

Permissionless - when the nodes' identity does not have to be known [Yeo+17], because all of them have the same permissions. In permissionless DLT designs with a large number of nodes (e.g. Ethereum), consensus finding is usually probabilistic and does not provide total finality, because it is impossible to reach finality in networks that allow nodes to arbitrarily join or leave. Consequently, the consistency between all the nodes of a public, permissionless distributed ledger can, at a certain point in time, only be assumed with a certain probability. Furthermore, a transaction appended to a distributed ledger is only assumed to be immutably stored to a certain probability. In blockchains, this probability of a particular transaction's immutability increases when new blocks are added to the blockchain [DL] [Sun20].

2.2.2 Properties and Characteristics

N. Kannengießer et al. [Kan+20] have extracted 277 DLT characteristics, which were eventually assigned to 40 master variables names and descriptions. These 40 resulting DLT characteristics (which we later discuss in more detail) were further grouped into 6 DLT properties:

Property	Characteristic	Description
Opaqueness	⋮	The degree to which the use and operation of a distributed ledger cannot be tracked
Performance	⋮	The accomplishment of a given task on a distributed ledger under efficient use of computing resources and time
Flexibility	⋮	The degrees of freedom in deploying applications on and customizing a distributed ledger
Security	⋮	The likelihood that functioning of the distributed ledger and stored data will not be compromised
Policy	⋮	The ability to guide and verify the correct operation of a distributed ledger
Practicality	⋮	The extent to which users of a distributed ledger can achieve their goals with respect to social and socio-technical constraints of everyday practice

2.3 Blockchain

The most commonly used data structure for distributed ledgers is the blockchain [ZJ18]. Each block contains a set of records added to the ledger and is immutable once generated. A newly created block is inserted in the blockchain by linking it to the last block in the chain. To prevent tampering of information found in existing blocks, the integrity of each block can be verified using a hash function which takes into consideration all preceding blocks. Hence, in order to successfully alter an older block, one must also modify all following blocks, which is considered unfeasible or unlikely. This implies that the amount of trust in the information contained in

a block depends on the block age (i.e., the number of blocks following it). Due to its strong guarantees, blockchains are the most prevalent form of storage for ledgers.

More recently, smart contracts have been introduced to support general applications beyond cryptocurrency. Smart contracts are programs automatically executed by the blockchain miners whenever their encoded conditions are triggered. Smart contracts are also transparent since they can be reviewed and agreed upon by the interacting parties prior to inserting them to the blockchain. Finally, the correct execution of smart contracts is guaranteed due to the immutability properties of a blockchain, which prevents it from being corrupted. Two notable systems which support smart contracts are Ethereum and Hyperledger.

The scope of blockchain applications has increased from virtual currencies to financial applications to the entire social realm. Based on its applications, blockchain is delimited to Blockchain 1.0, 2.0, and 3.0 [XCK19].

Blockchain 1.0 was related to virtual currencies, such as bitcoin [DL], which was not only the most widely used digital currency but it was also the first application of blockchain technology [MS15]. Blockchain 1.0 produced a great many applications, most of which were digital currencies and tended to be used commercially for small-value payments, foreign exchange, gambling, and money laundering. At this stage, blockchain technology was generally used as a cryptocurrency and for payment systems that relied on cryptocurrency ecosystems.

Blockchain 2.0 Broadly speaking, Blockchain 2.0 includes Bitcoin 2.0, smart-contracts, smart-property, decentralized applications (Dapps), decentralized autonomous organizations (DAOs), and decentralized autonomous corporations (DACs) [Swa15]. However, most people understand Blockchain 2.0 as applications in other areas of *finance*, where it is mainly used in securities trading, supply chain finance, banking instruments, payment clearing, anti-counterfeiting, establishing credit systems, and mutual insurance. The financial sector requires high levels of security and data integrity, and thus blockchain applications have some inherent advantages. The greatest contribution of Blockchain 2.0 was the idea of using smart-contracts to disrupt traditional currency and payment systems. Recently, the integration of blockchain and smart contract technology has become a popular research topic in problem resolution. For example, Ethereum, Codius, and Hyperledger have established programmable contract language and executable infrastructure to implement smart contracts.

Blockchain 3.0 is described as the application of blockchain in areas other than currency and finance, such as in *healthcare*, government, science, culture, and the arts [Swa15]. Blockchain 3.0 aims to popularize the technology, and it focuses on the regulation and governance of its decentralization in society. The scope of this type of blockchain and its potential applications suggests that blockchain technology is a moving target [Cro+16b]. Blockchain 3.0 envisions a more advanced form of "smart contracts" to establish a distributed organizational unit that makes and is subject to its own laws and which operates with a high degree of autonomy [Pie+18]. The integration of blockchain with tokens is an important combination of Blockchain 3.0. Tokens are proofs of digital rights, and blockchain tokens are widely recognized thanks to Ethereum and its ERC20 standard. Based on this standard, anyone can issue a custom token on Ethereum and this token can represent any right or value. Tokens refer to economic activities generated through the creation of encrypted tokens, which are principally but not exclusively

based on the ERC20 standard. Tokens can serve as a form of validation of any right, including personal identity, medical records, currency, receipts, keys, event tickets, rebate points, coupons, stocks, and bonds, etc. Consequently, tokens can validate virtually any right that exists within a society. "Blockchain is the back-end technology of the new era, while tokens are its front-end economic face. The combination of the two will bring about major societal transformation." [XCK19]

2.3.1 Hyperledger Fabric

Hyperledger Fabric [DMH17] is a distributed ledger platform for running chaincode (smart contract in Fabric). It is a specific blockchain platform, which is optimized for a specific task such as tracking assets, transferring values, etc. The modular architecture delivers high degrees of resiliency, flexibility, confidentiality, in design and implementation. The flexibility in design leads to achieving scalability, privacy, etc. Fabric is designed to support pluggable implementations of a different functions and chaincodes (using Go, Java, JavaScript). Transactions in Fabric are private and confidential thanks to its channelization features [Bal17]. Rather than an open, permissionless system, Fabric offers a scalable and secure platform that supports private transactions and confidential contracts. This architecture allows for solutions developed with Fabric to be adapted for any industry, thus ushering trust, transparency, and accountability for institutions [SSS18].

Other DLTs that were originally designed for ad-hoc, public use (where there is no privacy and no governance) had to be later significantly redesigned to add in support for permissions and privacy; Hyperledger Fabric was designed with these features as foundational. In this regard, Hyperledger Fabric has had a head start over many of the competing frameworks. For example, while there may be promise in some of the Ethereum 2.0 implementations, these are still mostly oriented to public network use [DMH17].

2.3.2 Ethereum

Ethereum [] is an open blockchain platform that allows anyone to build and use decentralized applications that run on blockchain technology. Financial interactions or exchanges could be carried out automatically and accurately using code running on Ethereum. It is a general purpose blockchain platform, which allows users to write their own algorithmic code and running customised logical processes. It was designed to be flexible and adaptable and has a powerful shared global infrastructure. The movement of assets around the network represents the ownership of property. In some ways, Ethereum is similar to that of Bitcoin, but there some technical differences between them. Bitcoin offers peer to peer electronic cash system, while Ethereum blockchain focuses on running the smart contract code of any decentralized application. Miners work to earn the crypto token Ether, this is also used to pay transaction fees and services in Ethereum network [SSS18].

The Ethereum network can be either public or private. Public DLT designs bring trust, security and transparency. Everything is recorded, public, and cannot be changed; also the more decentralized and active a public DLT design is, the more secure it becomes; and in terms of transparency - all data related to transactions is open to the public for verification.

3 Method

In order to answer our *RQ*, we conduct a literature review of the available studies on DP, DLT, Healthcare, Finance and DLT-based DP approaches for healthcare and finance. We consider a topic-centric [WW02] approach through a hermeneutic framework [BC14], that describes the literature review process as fundamentally a process of developing understanding that is iterative in nature. Using the hermeneutic circle it describes the literature review process as being constituted by literature searching, classifying and mapping, critical assessment, and argument development. The hermeneutic approach emphasizes continuous engagement with and gradual development of a body of literature during which increased understanding and insights are developed.

The process we have followed consists of two phases: *search and acquisition* and *analysis and interpretation*, which describe two circles, respectively, that are mutually intertwined. The first circle consists of *searching, sorting, selecting, aquiring, reading, indentifying and refinig* and is the first step of the second (bigger) circle, followed by the steps *mapping and classifying, critical assessment, argument deveopment, research problem/questions*. Through iterations on these circles we identified five stages, each of which poses a question (Q1-Q4, followed by the *RQ*), whose answer provides a mapping (or classification) that, together with critical assessment and argument development, serves as a means to go deeper into the subject by posing a follow-up question. Going multiple times through the two phases and through the processes of individual stages, our main aim is to get an overview of the emerged relationships. This will help us to eventually determine which of the DLT characteristics make them suitable for personal DP in healthcare and finance and why, if at all, thereby answering our *RQ*.

To make sure that the studies included in the review were clearly related to the research topic, we defined detailed general guidelines for inclusion and exclusion criteria for each of our five stages. Taking this into account, we will take a look the following five stages and the four corresponding questions leading to our *RQ*:

Q1: What are the fundamental requirements for a personal DP approach?

The scope of this stage is limited to the literature that (1) presents or describes solutions for research in DP systems within the computer science context, and/or (2) perform any type of quality analysis of these systems (surveys, taxonomies, ontologies, comparisons, categorisations) and/or (3) studies that discuss handling data of sensitive or private nature. Here we did not impose any restrictions on a specific domain of application.

Q2: What is the importance of individual DP requirements in terms of healthcare and finance?

By identifying specific personal DP requirements through the selected literature in the first stage, here we searched for studies that (1) research or discuss the importance or relevance of a requirement, and/or (2) literature directly researching DP, where any of these personal DP requirements are mentioned. This time we include only papers in the domain of healthcare

and finance.

Q3: Which are the preferred DLT approaches for DP in healthcare and finance and why?
The scope of this stage is limited to literature on (1) DLTs in the domain of (2) healthcare and finance.

Q4: Which requirements are considered important by DLT-based DP approaches in healthcare and finance?
The scope of this stage is limited to literature on (1) DLTs in the domain of (2) healthcare and finance.

RQ: What are the characteristics of DLTs that make them suitable for personal data provenance in healthcare and finance?
Finally, in stage five we make a selection of papers on DLTs, looking for those which specially tackled the (1) considered approaches or (2) use cases, or (3) perform any type of quality analysis of these systems (surveys, taxonomies, ontologies, comparisons, categorisations) or (4) discuss DLT characteristics. By mapping the results from the previous stages to these DLT characteristics, we aim to answer the proposed *RQ*.

The studies included in this work were identified through a thorough search for relevant published studies. Methods included conducting computer searches, "snowballing" procedures [SR09], examining relevant bibliographies, searching reference sections of the studies included in the relevant papers to identify further relevant studies, and contacting relevant researchers and organizations.

In table [..] we have summarised our selections of search strings based on some commonly used terms and acronyms for data provenance, individual requirements, healthcare, finance, etc. For example: "blockchain" OR "distributed ledger" AND "health" OR "review".

Using mainly Google Scholar, we found around 170 relevant studies. More than 50 are focused on DP and personal data, of which 8 discuss healthcare and 6 look at finance; another 30 papers in healthcare and 33 in finance helped identify important requirements for each use case respectively; there are 6 papers which compare and discuss different DLT approaches (2 in the form of taxonomies); around 50 studies investigate blockchain, where 21 are focused on applications in healthcare, 17 on applications in finance and 11 addressing features, differences and trade-offs between public and private approaches.

We excluded pure discussion or opinion papers, tutorials, and studies that tackles provenance in a context other than the computer science field. We also exclude studies reported in a language other than English. It is also important to point out that the literature search was conducted without any time restrictions, considering that these topics are relatively new, and therefore, all literature in those areas was considered relevant to our study.

	Data	AND	**Term**	OR	**Type**
OR	data provenance		health		survey
	data lineage		medical		review
	transparency		biomedical		systematic review
	tracing		clinical		literature review
	tracking		EHR		ontology
			EMR		taxonomy
	identifiability		PHR		case study
	linkability		patient		technical report
	anonymity		doctor		proof-of-concept
	pseudonymity		surgeon		report
	ownership		physician		
	access				
	scalability		finance		
	interoperability		money		
	security		bank		
	confidentiality		fintech		
	integrity		defi		
	availability		payment		
	traceability		consumer		
	trust		customer		
	policies				
	law				
	regulation				
	usability				
	ease of use				
	blockchain				
	distributed ledger				
	hyperledger fabric				
	ethereum				

4 Results

4.1 Requirements

Q1: What are the fundamental requirements for a personal DP approach?

In recent years there has been a rapid growth of the provenance field, in general, and of DP solutions, in particular, which has derived into a large and heterogeneous research corpus of approaches to address a variety of DP concerns. Even so, at the present time there appears to be no clear consensus or common ground on aspects such as what requirements (a necessary feature) a DP approach should support or what technical details are involved in making these systems possible.

By analysing the relevant literature we found on DP, we aim to answer *Q1*, and, thereby, determine and identify the most fundamental requirements that a DP approach, suitable for tracing personal data, has to fulfil.

We put these requirements into three groups to provide context: "Data Subject" consists of requirements important to the Data Subject (**Identifiability, Ownership, Accessibility**), whereas "System" contains requirements for the particular tool/approach (**Scalability, Interoperability, Security, Traceability, Trust**). "Other" includes two requirements necessary in almost every system, however, we think they are still important to consider, because of this works' particular focus on the average user and their personal data.

It is important to note that, for example, **Identifiability** encompases the concept of identification, as well as pseudonymity, anonymity and unlinkability. **Security** is defined by confidentiality, integrity and availability (CIA). Also, these requirements influence each other and while sometimes equally important, they are often incompatible with one another. For example, by striving for **Scalability** and/or **Interoperability**, one must sacrifice Security, thereby potentially damaging consumers' **Trust** in the system. Also, for example, **Trust** is positively influenced by other secondary requirements such as accountability, system auditability, durability, data completeness, granularity, consistency, verifiability, authenticity, as well as the fulfilment of other requirements in this table (e.g. **Security**, **Traceability**, etc.). The following table presents the requirements we identified as fundamental for personal DP:

Group	Requirement	Description
Data Subject	Identifiability	An unique identifier allows identification and lays the ground for accountability [Lee+13]. However anonymity, pseudonymity and unlinkability are as important. [HPH11; Sen].
	Ownership	Allows Data Subjects to get an overview, request or perform changes and deletion of the data that they own. [ZN+15]
	Accessibility	Allows Data Subjects with access to view, store, retrieve, move or manipulate data, based on their access rights [ZN+15; BKB16].
System	Scalability	With the increase of the data volume and the number of operations, it should be possible to store and process provenance information efficiently and without risk of information loss [TBA16; Fre+08, p. 16].
	Interoperability	The ability of different information systems, devices or applications to connect, in a coordinated manner, within and across organizational boundaries to access, exchange and cooperatively use data amongst stakeholders [, IntOp].
	Security	Ensures non-disclosure of data traveling over the network to unauthorised Data Subjects (confidentiality) [Asg+12]. Ensures that the Data Receiver may detect unauthorised changes made to the data (integirty) [Tsa+07]. Ensuring that data and is provenance is available to Data Subjects,when and where they need it (availability) [Lia+17].
	Traceability	In this work we consider traceability and transparency synonymous to each other. It means providing information on what transmitting principle was used, what type of data, for what purpose and to whom the information was sent. How data is collected; how, when, where it is stored [Fre+08; ZN+15, p. 13].
	Trust	If the Data Subject trusts the system, they seem to be willing to share personal information [BHS02]. The willingness to share data can also increase if the Data Subject finds the advantages of engaging in such a transaction more valuable than the loss of privacy [BGS05; AG05].
Other	Compliance	Enforcing laws [], policies and regulations such as purpose limitation [FHS17], data minimisation [ASS17], etc.
	Usability	Provides clear interfaces and structures that display provenance information in an understandable way (usage of icons, graphs, etc.). Managing security (and privacy) is not the primary task of the user [Fre+08].

4.2 Use Cases

Q2: What is the importance of individual DP requirements in terms of healthcare and finance?

In this work we investigate DP technologies for both *healthcare* and *finance* and in order to answer *Q2*, we take a closer look at each individual DP requirement. While all of the above mentioned requirements are important in such technologies, with this approach, we aim to develop an overview and understanding of the nuances and differences between both use cases, in terms of the DP requirements. This is an important step, in order to accurately compare and map them to our considered approaches' features further in the work and have a basis for evaluation of the requirements fulfilled by DLT-based DP approaches in *healthcare* and *finance*. By analysing how each DP requirement was expressed through the lens of *healthcare* and *finance* respectively, we can observe the differences that emerge due to the nature of the use cases.

4.2.1 Healthcare

Actors: *Patient, Physician, Institution*

In regard to medical treatment and patient safety, the importance of data, its origins and quality have long been recognised in clinical research [Cur+17] [Muh14]. Creating trust relationships among the various actors is vital - e.g., evidence-based medicine and healthcare-related decisions using third-party data are essential to patient safety [Mar+20]. DP is also crucial for solving confidentiality issues with healthcare information like accidental disclosures, insider curiosity and insider subornation [Rin97b]. In the following we discuss the important aspects of each of the DP requirements pointed out in table 2.3 in terms of healthcare.

Identifiability: There are important trade-offs between indentifiability and unlinkability/anonymity. For example, a patient feels that their physician misrepresented a test and wants to share this information, but is reluctant to do so, since casting the physician in a negative light can have repercussions in their care at a later time []. Another example is the perceived stigma of having a mental disorder acts as a barrier to help seeking. It is possible that patients may be reluctant to admit to symptoms suggestive of poor mental health when such data can be linked to them, even if their personal information is only used to help them access further care. There is a significant effect on reporting sub-threshold and non common mental disorders when using an anonymous compared to identifiable questionnaire [Fea+12]. Studies suggest that anonymity is strategically used and fosters self-disclosure among individuals who are embarrassed by their illness [Rai14].

On the other hand most people believe that, when a physician makes an error, an incident report should be written and the individual should be identified on the report. People are reluctant to accept physician anonymity, even though this may encourage reporting [Eva+04]. Also, Data Protection Act insists that patients must consent directly to participate in research or that patients' data must be completely anonymised. However, this causes particular problems for epidemiological research [War+04] which often requires access to routinely collected identifiable personal data, or requires identification of research participants from such data. Obtaining

individual consent from large numbers of patients may be onerous or simply impossible, for example if patients have died or moved away, and participation bias may undermine the data. Anonymising data is difficult and expensive and greatly limits their future value [Wal06].

Ownership: A relevant issue is the ongoing debate about the ownership of patient data among various stakeholders in the healthcare system including providers, patients, insurance companies and software vendors. In general, the current model is such that the patient owns his/her data, and the provider stores the data with proprietary software systems. The business models of most traditional EHR (electronic health record) companies are based on building proprietary software systems to manage the data for insurance compensation and care delivery purposes. Such approach does not encourage or makes it difficult for individual patients to share data for scientific research, nor does it encourage patients to obtain their own health records that may help better manage their health and improve patient engagement [Adi+17].

Accessibility: It is important that the different actors can view, store, retrieve, move, request changes/deletion or manipulate medical data based on their access rights [Ber17]. For example, patients should be able to see what prescriptions they have so they know what medicine to take; physicians should be able to alter the prescriptions of their patients and also to see what prescription a patient has gotten from other physicians so that they can correctly treat them and avoid medication errors; an institution should be able to verify a patient's prescription to make sure that they are not trying to purchase unintended pharmaceuticals [, Priv].

Scalability: The amount of global healthcare data is expected to increase dramatically by the year 2020. Early estimates from 2013 suggest that there were about 153 exabytes (10^{18}) of healthcare data generated in that year. However, projections indicate that there could be as much as 2,314 exabytes of new data generated in 2020 (around 15 times more) []. However, because of the sensitive and private nature of healthcare data, it is important that scalability should not come at the expense of security and trust.

Interoperability: Healthcare is considered as a domain with growing focus on interoperability. Products obeying international standards will improve quality and sharing . Interoperability helps policy makers and project coordinators in defining long term strategies by providing software sustainability and securing the investments. Moreover, interoperability enforces security and patient safety: the quality of the patient healthcare treatment is not depending on the quality of a specific software solution (the so-called vendor lock-in effect). Using international standards forces vendors to comply with the state-of-the-art of the security measures [Mar+20]. This interoperability is a fundamental requirement for the health care system to derive the societal benefits promised by the adoption of electronic medical records (EMRs). One critical question is whether the adoption of EMRs needs to wait for interoperability standards or whether it can proceed efficiently without them [Bra05].

Security: Confidentiality and trust between a physician and patient is not new: it is central to the practice of healthcare and has been focused on since Hippocrates. Whilst the concept of patient confidentiality has endured as an ideal throughout history. In the digital age, patient confidentiality is often framed within the context of electronic patient records and the potential

involvement of third parties. While the involvement of institutions and other research organisations can resolve many practical issues for healthcare providers, it often involves the transfer of sensitive patient information to these institutions [Rin97b]. Therefore, it is important that there isn't any disclosure of medical data traveling over the network to unauthorised actors [Rin97a, p. 96]. Sometimes, however, difficulties with keeping the confidentiality of personal health information may arise, because of the often unclear position of family members and friends, in patient's health and medical treatment [Pet+04].

Data Integrity issue is one of the most demanding concerns for the healthcare industry in the whole world. An integrity breach in a healthcare organization can have disastrous consequences. A patient whose data has been tampered with could be given wrong medications causing fatalities. Most healthcare organizations at present have weak and vulnerable data storage procedures and lack secure mechanisms to foil malware attacks. All these issues create many challenges associated with data integrity in healthcare organizations [Pan+20]. A more concrete example is preserving the integrity of medical images through watermarking schemes [BC16]. Medical images transmitted through the network can be easily tampered and forged, which increases the risks of misdiagnosis. Therefore, the image authenticity and integrity have become two crucial security factors in e-Health applications [Liu+19].

The immediate Availability of patient and resource oriented information is of great importance, in order for physicians and institutions to, for example, identify the most appropriate ambulance and healthcare setting; provide guidance to physicians as to the most appropriate management of the emergency case at hand; prioritize/classify the emergency case and overall improve the quality of the emergency care [PMV12]. Medical data and its provenance should be available and ready for immediate use, especially in cases of emergency [KLG03].

Traceability: Traceability in healthcare is at the crossroads of numerous needs. It is therefore of particular complexity and raises many new challenges. Identification management and entity tracking, from serialization of pharmaceuticals, to the identification of patients, physicians, locations and processes is a huge effort, tackling economical, political, ethical and technical challenges. There are growing needs to increase traceability for drug products, related to drug safety and counterfeited drugs [KS18]. Technical problems around reliability, robustness and efficiency of carriers are still to be resolved. Traceability is a major aspect of the future in healthcare and requires the attention of the community of medical informatics [Lov08].

Trust: Trust is, of course, essential to both physician and patient. Without trust, it is difficult for a physician to expect patients to reveal the full extent of their medically relevant history, expose themselves to the physical exam, or act on recommendations for tests or treatments [Saf+98; Mos+98]. Trust promotes efficient use of both the patient's and the physician's time. Without trust, the process of informed consent for the most minor of interventions, even a prescribed antibiotic, would become as time consuming as that needed for major surgery [Goo02]. Furthermore, physician-to-patient relationship is jeopardised when people do not trust that their personal health information will be kept confidential, and that these data will not be utilised for purposes other than medical [KLG03].

It is also suggested that it is morally important for doctors to trust patients. Doctors' trust of patients lays the foundation for medical relationships which support the exercise of patient autonomy, and which lead to an enriched understanding of patients' interests. It may not be

possible to trust at will, the conscious adoption of a trusting stance is necessary as the burdens of misplaced trust fall more heavily upon patients than physicians [Rog02]. In terms of medical research, one of the three key factors to the patients willingness to share data is contingent upon trust who is accessing the data [KMR19].

Comliance: Unfortunately, legal controls over data collection in European countries have badly affected the work of epidemiologists [WN94a]. While data protection laws, policies and regulations aim to protect the patients information, rights and health, they might cause harm to the patients well-being in the long run, by damaging the ability of institutions to conduct unbiased and reliable medical research [War+04; WN94b].

Usability: For example, 30% of electronic medical record (EMR) system implementations fail, often because physicians cannot use them efficiently. User experience problems are wide-spread among EMRs. These include loss of productivity and steep learning curves [SMK09]. There is an increasing awareness of the need for higher usability of medical technology. This requires an understanding of what usability is and what usability evaluation methods are suitable, both in the design process and when medical technology is purchased at hospitals [Lil06]. Also, a big challenge is the lack of patient engagement in healthcare (not all patients are willing and able to manage their own data), which can be potentially improved through higher ease of use.

4.2.2 Finance

Actors: *Consumer, Institution*

In online banking, digital money and digital financial services, the importance of information about transactions, money flow, money origin, credit scores and financial decisions is becoming bigger and bigger since the emergence of e-finance [AHS02]. DP is of great use not only in investigating money laundering [Ung+06], tracing donations [Sir+19], charities [Sin+20] or illegal funding [Tei18], but also loans and financing, mortgages, trading of currencies, insurance policies and others [But20]. However, 'big tech' are also venturing into financial services [Boi+21]. While being accused for abuse of market power and anti-competitive behaviour, they are also famous for not giving extensive information on how personal data is analysed, processed or interacted with by third parties and international or government organisations [, RV19], which has a negative impact on the consumers' ability to trace their personal data.

Identifiability: On one hand, in the last ten years there has been a tendency to introduce anonymity into stock, bond, and foreign exchange markets. Almost all the asset markets organized as electronic platforms are anonymous []. On the other, The last few years have seen an international campaign to ensure that the world's financial and banking systems are "transparent," meaning that every actor and transaction within the system can be traced to a discrete, identifiable individual []. Anonymity fosters crime, while identifiability challenges privacy. For example, there is a high degree of anonymity with Bitcoin [DL], however traceability is possible [RH13]. In connection to this, a study shows that the relationship between participants' views on anonymity and traceability as a disadvantage to bitcoin transactions was

statistically significant [AW21]. Perhaps consumers should be able to perform operations in an pseudonymous way, that ensure ownership (pseudonyms are not improperly used by others) and ensure individuals are held accountable for abuses created under any of their pseudonyms [Cha85].

Accessibility: Not all information in e-finance is private. Indeed, by law, many types of transactions must be made available to various institutions, ranging from the government to the public. As a practical matter, there will often be several parties to a transaction who must have access to the information [SWR97]. Another example is money inheritance, where an institution or another consumer can give access rights to their personal financial data or money. This can mean that consumers require and can attain access rights to other consumers' or institutions' financial data.

Scalability: E-Finance is a constantly growing field. As of November 2021, there were over 10 000 fintech (financial technology) startups in the Americans, making it the region with the most fintech startups globally. In comparison, there were over 9 000 such startups in the EMEA region (Europe, the Middle East, and Africa) and over 6 000 in the Asia Pacific region. 25 000 new startups in 2021 compared to only around 12 000 in 2018 []. However, with big tech providing basic financial services through their low cost structures (especially where a large part of the population remains unbanked) [Boi+21], lagre-scale DP approaches should consider measures against discrimination, abuse of market power, anti-competitive and monopolistic use of data.

Interoperability: Without interoperability, consumers need to visit multiple institutions and systems to make transactions with different networks, which are subject to fees. If networks are interconnected, fees are expected to be lower. Thus, transactions are cheaper and more other consumers can be reached, which will increase the number of transactions [HB17].

Security: According to a study examining the conflict between anti-money laundering and anti-terrorism finance law requirements and bank secrecy and confidentiality laws [], the duty of confidentiality is regarded as an essential feature of the institution-consumer relationship and it was enunciated at a time when crime was viewed as a local phenomenon. However, the last two decades have seen the rise of transnational crimes such as money laundering [Ung+06] and terrorist financing [Tei18]. To counter these crimes a number of legislations were enacted which, require institutions to disclose their consumers' financial information in certain circumstances to law enforcement authorities. This is justified by the fact that institutions are used by criminals to launder criminal proceeds and the audit trail they leave behind helps criminal investigation and prosecution. However, this is still personal financial information and there exist the requirement for some level of confidentiality [Ber14]. Also, Integrity is important in finance, helping to generate the trust that is vital for a financial system to flourish [Cow02; SVK12; Boa11].

Traceability: The term traceability may have a law enforcement implication suggesting, for example, the ability to monitor or track the activies of consumers. While transaction recors and audit trails certainly can provide such a capability, this is different from using traceability to

verify the accuracy of a measurement or the authenticity of a set of data [SWR97]. Traceability can discourage fraud, and criminal activities like money laundering [Ung+06], illegal funding [Tei18] or simply bring transparency in donation tracing [Sin+20; Liu+21; SAD19]. There is also a research that supports the notion that transparency is a desirable characteristic of financial reports - increased transparency reduces information risk and cost of capital [BS08].

Trust: Financial transactions, being all exchanges of money over time, should be particularly dependent on trust. In fact any financial transaction, being it a loan, a purchase of a stock of a listed company or the purchase of an insurance policy, has a fundamental characteristic: it is an exchange of money today against a promise of (more) money in the future. But what leads the consumer to believe that promise and make the exchange actually possible, is trust. The trust of a consumer who has invested in the stock of a company that his money will not be appropriated by the company's managers [Gui12]. Currently trust in finance is highly dependant on third parties and intermediaries [JSZ07; Bos01], which also has its risks [LJ09]. Also, security and privacy, usability and reputation have a direct and significant effect on consumer trust in a financial services. Besides this, consumer trust is positively related to relationship commitment. It's also observed that trust is a key mediating factor in the development of relationship commitment in the online banking context [CFG07a].

Compliance: Since financial applications and services carry quite sensitive consumer personal data, there should exist a policy framework that provides comprehensive set of policies that aim to ensure security, transparency and trust [Hus+21]. Financial regulations can also stabilize the financial market and increase the benefit to consumers by promoting innovation and competition in the market [Yan17].

Usability: As the diversity of services in the financial market increases, it is critical to design usable tools in order to overcome the complex structure of the system [CFG07b]. Consumers are heavily encouraged to perform critical financial operations online, despite the continuing absence of appropriate tools to do so. Many security requirements are too difficult for consumers to follow, and some marketing-related messages about safety and security can actually mislead consumers [MV08]. Also, usability was found to have a positive effect on consumer satisfaction and satisfaction with previous interactions with the system had a positive effect on both consumer loyalty and trust [CFG08].

4.2.3 Comparison

To gain understanding of some of the differences, we can take a look at, for example, **Accessibility**. In both use cases some data should be accessed only by authorised individuals or institutions. However, in *healthcare*, due to the prominent hierarchical structure (patient, physician, institution), patients having access to the physician's or institutions' information, as well as patients sharing their health information between each other is uncommon. Physicians and institutions should operate with patients' personal health data with confidentiality. There is a stronger need for individual **Accessibility** control and access rights in comparison to *finance*, where, for example, many types of transaction and financial reports must be made available to various institutions, ranging from the government to the public.

In *healthcare*, **Scalability** must account for the bigger and bigger quantities of medical data generated, whereas for *finance*, it is suggested that such approaches should be able to handle the constantly growing number of participants.

Interoperability in *finance* means faster transactions with lower transaction fees. On the other hand, **Interoperability** in *healthcare* means potentially better treatment and patient safety. Additionally, medical data and its provenance should be always available and ready for immediate use, especially in cases of emergency, compared to *finance*, where the patients life and health are generally not dependant on the immediate availability of information.

Another important mention is the legal control over data collection in European countries (**Compliance)**. Laws, regulations and policy frameworks in *finance* only seem to benefit the consumer by ensuring security, transparency and trust. In *healthcare*, however, although aiming to protect patient's information rights and health, the same approaches can damage the ability of epidemiologists and institutions to conduct unbiased and reliable medical research, thereby, paradoxically, cause harm to the patients well-being in the long run.

4.3 Considered Approaches

Q3: Which are the preferred DLT approaches for DP in healthcare and finance and why?

Many of the existing DP approaches are based on a centralized storage model. The downside to the centralized system architecture is that if the central server is compromised, the whole DP trails could be compromised. According to the literature [RK+17], most DP systems do not try to validate the changes before they are stored. Blockchain or DLT in general, can serve as a medium for storing information and providing validations for each of the changes before logging the changes (using smart contracts, for example). The immutable nature of the blockchain environment ensure that the approved provenance changes cannot be modified by any users once they are stored. Due to the distributed nature of the blockchain, the DP trails can be replicated on every node of the blockchain ensuring high availability and fault tolerance.

Removing the need for a centralized trusted third party in distributed applications is, perhaps, the most obvious and outstanding benefit of blockchain. By making it possible for two or more parties to carry out transactions in a distributed environment without a centralized authority, blockchain overcomes the problem of single point of failure that we mentioned, which a central authority would otherwise introduce. It also improves transaction speed, by removing the delay introduced by the central authority, and at the same time, it makes transactions cheaper since the transaction fees charged by the central authority is removed. In place of a central authority, blockchain uses a consensus mechanism to reconcile discrepancies between nodes in a distributed application.

Another DLTs property is DP. The data storage process in any distributed ledger is facilitated by means of a mechanism called transaction. Every transaction needs to be digitally signed using public key cryptography (PKI) which ensures the authenticity of the source of data. Combining this with the immutability and irreversibility properties of a distributed ledger provides a strong non-repudiation instrument for any data in the ledger [Cho+19].

One example of DLT, namely blockchain, was initially employed as the public transaction ledger for cryptocurrencies. However, beyond cryptocurrencies, blockchain technology has been recently considered for a plethora of other applications [Mau+17; Mar+20; Mun+06] as it encapsulates unique characteristics including decentralization, security, transparency and provenance. Such characteristics are particularly advantageous for variety of prominent issues experienced in the financial sector. As a result, blockchain technology holds the potential to revolutionize the financial industry by altering the way in which different services are conducted. Healthcare is also a domain in which blockchain is expected to have significant impact. Research in this area is relatively new but growing rapidly; so, health informatics researchers and practitioners are always struggling to keep pace with research progress in this area.

After a detailed analysis of the available literature, we noticed that many studies seem to either use or suggest permissioned instead of permissionless DLT designs for the implementation of DP solutions in the healthcare and financial domain, due to the private and sensitive nature of the information [Has+20], as well as the more flexible, reliable and efficient nature of permissioned DLT designs.

In healthcare, the majority of applications are developed on the popular blockchain networks, such as Ethereum and Hyperledger Fabric [AME19; Has+20; Dag+18; HE18; HK21; Ram+18; Rah+20; Mar+20]. The leading use cases are health records (EMRs, EHRs, PHRs), followed by other blockchain use cases, such as the management of the drug/pharmaceutical supply chain, biomedical research and education, health insurance claim processing and remote patient monitoring [Höl+18; AME19].

Finance being a broad domain, numerous different DLT and blockchain approaches can be considered. It is not feasible to investigate them all in this paper, however, there are studies that consider Hyperledger Fabric as a blockchain framework in the financial industry [Bet+19; CTF18; AI19; Ma+19] and Ethereum is a general purpose blockchain platform, which is widely used for numerous financial services []. Some of the use cases we identified include payments, trade finance, credit scoring, donation tracing, loans and financing, mortages, insurance policies, trading of currencies, etc.

The high level of consideration of these approaches in healthcare and finance research serves as motivation to focus our attention on these two DLT designs: Hyperledger Fabric and Ethereum, consequently answering *Q3*.

In the following two subsections we aim to map Hyperledger Fabric's and Ethereum's characteristics to our DP requirements.

4.3.1 Hyperledger Fabric

In Hyperledger Fabric, since the network is permissioned, every user participating in a transaction must register in the network for getting their corresponding IDs, which can ensure high **Identifiability** [SSS18]. The identity of each entity within the network is verified, thereby enabling auditability and accountability. However, high **Identifiability** is not necessarily a good feature in terms of DP in healthcare and finance. As previously mentioned, while physician accountability in case of error is desirable, data anonymity is as important in reporting mental illness and conducting medical reasearch; in finance anonymity fosters crime, while

Identifiability challenges privacy. There is also a tendency to introduce anonymity into stock, while there are campaigns to ensure transparency and **Identifiability** in banking systems.

Another one of the many compelling Fabric features is the enablement of a network of networks. Members of a network work together, but because agents need some of their data to remain private, they often maintain separate relationships within the networks [DMH17]. The modular architecture of Fabric helps to achieve data confidentiality, **Scalability** and **Security**. Incorporating chaincodes (smart contracts) can ensure control over authorizations, **Accessibility** privileges and data **Ownership** on the blockchain network, which we consider essential for DP (req table).

Fabric's high throughput (3000 transactions per second) and low latency relate to easier **Scalability**. Here it is important to note that, while Fabric can scale in terms of number of transactions per second, it's not as easily scalable in terms of new actors constantly joining the network (due to its private, permissioned nature). This makes Fabric a potentially better candidate for a healthcare DP solution (given the constantly growing volume of clinical data involved), compared to finance, due to their previously mentioned differences in terms of **Scalability**.

High degree of **Interoperability** can also be observed thanks to its multi-language smart-contract support (Go, Java, Javascript) and Fabic's support for EVM and Solidity. Fabric also claims to achieve high level of **Security** thanks to its robustness, private transactions, confidential contracts, data isolation, governance of smart contracts, etc. However, due to lack of international standards in healthcare and the overall high number of different DLT designs used in healthcare and finance, **Interoperability** remains one of the most difficult challanges. Applications developed by different vendors or on different platforms may not be able to interoperate. For example, two systems, where one is developed on the Ethereum platform while the other is developed on the Hyperledger Fabric platform makes it challenging to exchange information from one platform to the other.

Access rights, linkability through identification and the nature of blockchain can provide high degree of **Traceability**. Its highly modular and permissioned architecture can bring confidentiality and **Security**. This, combined with blockchain's property of decentralization, immutability, data provenance, auditability, accountability, reliability, etc. can help achieve **Trust** and make it suitable for storage and management of patients' health records or consumers' financial information.

However, the credibility of a private network relies on the credibility of the authorized nodes, which means they need to be trustworthy as they are verifying and validating transactions. In terms of **Usability**, the upgradability of chaincode in Fabric is vital when new logic needs to be incorporated or a bug needs to be fixed and queryable data is also a feature which can positively impact ease of use. [DMH17].

4.3.2 Ethereum

Ethereum (public), being a public ledger, has strong support for auditability and accountability in case the proper identity of entities can be verified and it provides pseudonymous **Identifiability** via public key, which, as we previously mentioned, can be beneficial for DP in both healthcare and finance. Ethereum executes random complexity codes through its EVM and is freely available without authorization to any person. Due to the fact that Ethereum is totally outside

of every particular area of application and serves more as a general interface for all types of transactions and applications. **Ownership** and **Accessibility** are not necessarily provided features, nevertheless they can still be implemented through smart contracts.

Furthermore, Ethereum lacks speed and faces concerns over **Scalability** (due to it's probabilistic finality [Vuk15]), energy consumption and cost [Cho+19]. The bigger the public network, the slower it is, as more transactions take place and clog the network. From the **Scalability** perspective, Blockchain-based systems like Ethereum, are commonly compared [Cro+16a] with conventional systems like VISA (2000-56000 transactions per second vs. 13 of Ethereum). This quantitative comparison shows that Ethereum is quite far from offering a viable implementation platform for all transaction-based systems [BFV19]. Data stored in these ledgers is also visible to any participant and therefore not suitable to handle sensitive data, which on the other hand benefits transparency and general DP. Storing data also incurs expense and is therefore infeasible to store a large amount of data in the ledger. On the other hand, due to the network's public nature, new actors can easily join the network and take part, without requiring permission. This can be really beneficial in terms of **Scalability** and the contantly icreasing number of fintech startups and unbanked individuals entering the domain through new (blockchain) approaches for financial services.

According to the scalability trilemma [BFV19], at the expense of **Scalability**, we get **Security** and decentralisation (**Trust**). Ethereum's immutable computer logic, together with smart contracts, EVM, consensus mechanism, Ether (token, used to store, process and update data via transactions) and lack of necessity for trusted third parties, further increase the **Traceability** of information and **Trust** in the system. The immutability and irreversibility of the ledger can provide integrity. In terms of availability, the public Ethereum network is best suited to storing long-term static data that need to be widely available, such as the Ethereum Registration Authority information [Rob20] (**Security**), which makes it particularly useful for patients' health records or consumers' financial information.

Thanks to its customisability through Turing-complete smart contracts, its generalized purpose, flexibility, adapatability, current success and available tools, research and information for developers and users, it can be seen as accessible, relatively widespread and easy to use/setup (**Usability**). However, smart contracts in Ethereum lack any upgradability feature, thereby making it difficult to update any smart contract in case a new feature needs to be added or a bug needs to be corrected. Ethereum provides also the posibility of decentralised applications (DApps) and decentralised autonomous organisations (DAOs), which can create or enable users to vote for policies and regulations (**Compliance**).

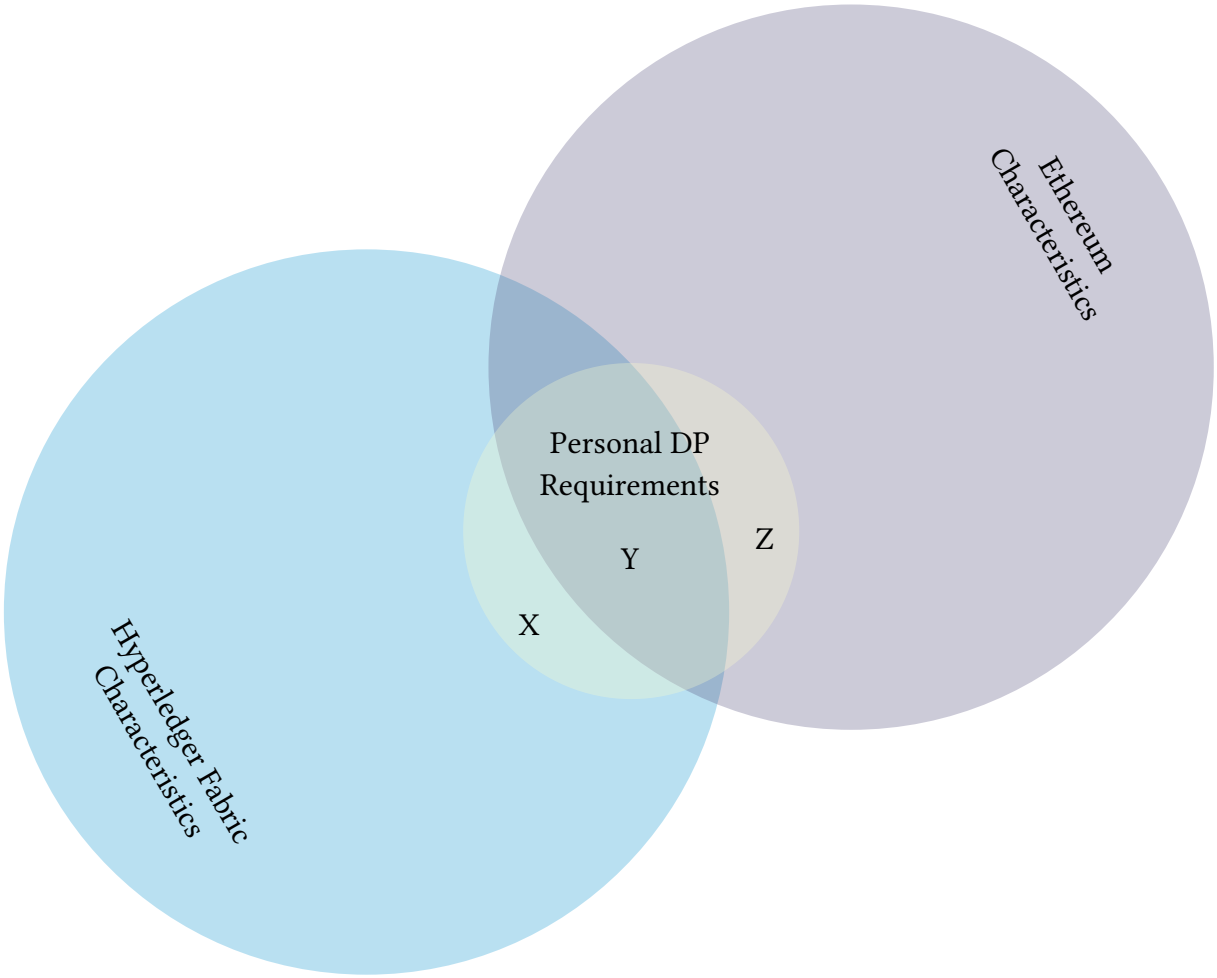
Ethereum can also be used as a private network. Approaches like sharding [], side-chains [21a] or choosing non-default values for the chain parameters can affect throughput, latency and drastically increase **Scalability** [SDS19]. Private Ethereum sidechains promise also to ensure confidentiality [JRB19]. While researchers and developers recognise it as challenging and complex, some solutions to cross-chain **Interoperability** have been proposed by a number of private Ethereum approaches [; 21b; Cle19]. On the other hand, the probability for successful partition-based attacks [NG17] increases in private, forkable DLT designs such as a private Ethereum blockchain, which increases the likelihood for violations of a distributed ledger's immutability. Increased vulnerability for immutability violations reduces the integrity of a distributed ledger [Kan+20], which can affect the system's level of **Security** and damage consumers' **Trust**.

4.3.3 Comparison

The following table provides an overview of each considered aproach's degree of fulfilment of our DP requirements and serves as a summary of the detailed description from the previous two subsections.

DP Requirement	***Hyperledger Fabric***	***Ethereum (public)***	***Ethereum (private)***
Identifiability	high	medium	high
Ownership	low	low	low
Accessibility	medium	low	medium
Scalability	medium	low	medium
Interoperability	high	low	medium
Security	high	high	medium
Traceability	high	high	high
Trust	medium	high	medium
Compliance	medium	high	high
Usability	high	medium	medium

Additionally, the following diagram displays where the personal DP requirements are positioned in terms of their fulfilment by the considered approaches. This doesn't mean that Hyperledger cannot achieve Compliance, or there aren't any proposed solutions to Ethereum's Scalability and Interoperability problems. It means that the personal DP requirement can be fulfilled to a higher degree by Hyperledger Fabric (in X) and Ethereum (in Z), respectively.



X	Y	Z
Scalability	Identifiability	Compliance
Interoperability	Ownership	Trust
	Security	
	Traceability	
	Usability	

4.4 Considered Requirements

Q4: Which requirements are considered important by DLT-based DP approaches in healthcare and finance?

In order to answer the question we will take a look at which requirements are considered important by the available literature on DLT in healthcare and finance, namely, what benefits do DLTs bring and also what challenges remain.

4.4.1 Healthcare

The healthcare sector is a problem-driven, data- and personnel-intensive domain where the ability to access, edit and trust the data emerging from its activities are critical for the operations of the sector as a whole [Has+20]. The activities of health institutions are tightly interwoven and require effective interchange of consents, patient-related data and proofs, and reimbursements processes, which effectively means exchanging data between different institutions. At the same time, health institutions are mandated to protect the highly sensitive data that patients choose to share with them.

4.4.1.1 Overview

In a detailed review of blockchain solutions in healthcare [Has+20], it is suggested that to both maintain the patient's privacy and exchange data with other institutions in the healthcare ecosystem, accessibility control, DP, data integrity and interoperability are crucial requirements. The traditional way of achieving access control commonly assumes trust between the owner of the data and the entities storing them. These entities are often servers fully entrusted for defining and enforcing access control policies. Interoperability is necessary for coordination, cooperation and optimisation of the health of patients. DP can deliver auditability, traceability and transparency in EHR, and help achieve trust in EHR software system. The authors argue that many of the currently existing issues in healthcare like unauthorised sharing, robbery of sensitive data, malpractices within the healthcare ecosystem, overall exploitation of trust, counterfeit drugs, etc., can be improved through blockchain.

In the same study, the processes within the targeted blockchian systems were mostly focused on sharing, storage, exchange and access of medical data, aligning with our definition of DP, thereby providing excellent source of knowledge as to which of our DP requirements are considered important in the available literature on DLT in healthcare. The study identified that 35% of the reviewed publications focused on impoving access control, 27% discussed solutions for the interoperability challenges, and 12% targeted the ability to improve DP overall and 28% of the included publications proposed increasing data integrity just by benefiting of the blockchain's key characteristic of immutability.

It is important to note, that, where defined, the studies mostly consider permissioned designs. Public persmissioned blockchains appear to be the preferred design choice, since the healthcare domain deals with highly sensitive data, which usually entails that a limited number of entities should have access, a public permissioned blockchain may be more appropriate than a public

permissionless and private, in order to ensure that data is not accessible by those who have no view rights and also to comply with current health and personal data regulations.

As mentioned in the previous section, in all of the reviews that we analysed, the majority of solutions are based on Ethereum, followed by the Hyperledger Fabric framework. The utilization of smart-contracts (or chaincodes in Fabric) partly explains why these are the mostly used platform for the proposed concepts. A smart-contract function, which often has the purpose of reducing third party interaction, has the potential of making health informatic processes more efficient. This choice of platform/framework correlates well with the overall popularity of blockchain platforms. The reasons for this can be both the attributes that are offered by the respective platform, but also the number of developers available with knowledge on each platform as well as the strong overall market position of Ethereum and Hyperledger.

Suprisingly, the most commonly used consensus algorithm turned out to be proof-of-work (PoW) [Höl+18]. Healthcare is an environment where the speed of transactions could be very important and PoW is a very slow consensus algorithm. PoW is also generally not used with permissioned chains where the actors are generally known (in this case, patients, physicians, institutions) and trust is easier to establish. Permissioned chains are usually more centralized (e.g., hospitals) than public networks where anybody can be a node and where PoW is the dominant consensus algorithm. The PoW is also a very computationally demanding algorithm. It would be impractical if hospitals would have to establish large computer centres just to mine the transactions. However, Ethereum may strengthen its position further as a blockchain network solution for healthcare in the future, since it is said to make a transition into proof-of-stake, which as a consensus algorithm, seems to fit around most of the discussed PoW limitations.

4.4.1.2 Benefits

Another comprehensive literature review [AME19] suggests that blockchain characteristics (or DLT in general) like decentralisation, improved data security, data ownership, availability, trust and data verifiability are clearly beneficial in healthcare applications.

"Blockchain can become that decentralised health data management backbone form where all the stakeholders can have controlled access to the same health records, without anyone playing the role of central authority over the global health data"[AME19, p.7]. The fact that the information in the blockchain is replicated among all the nodes in the network creates an atmosphere of traceability and openness, allowing healthcare stakeholders, and in particular the patients, to own their data and be in control of how their data is used, by whom, when and how, thereby enabling data provenance.

Additionally, compromising any one node in the blockchain network does not affect the state of the ledger since the information in the ledger is replicated among multiple nodes in the network. Therefore, by its nature, blockchain can protect healthcare data from potential data loss, corruption or security attacks. Also, the immutability property of blockchain which makes it impossible to alter or modify any record that has been appended to the blockchain aligns very well with the requirements for storing, sharing and tracing access of healthcare records.

Furthermore, since the identities of the patients in a blockchain are pseudonymized through the use of cryptographic keys, the health data of patients may be shared among healthcare

stakeholders without revealing the identities of the patients. In many studies smart contracts are also considered a feature, which can be used to program the rules that allow the patients to be in control of how their health records are shared or used. This is particularly relevant to the European General Data Protection Regulation (GDPR) which prohibits the processing of sensitive personal data of patients unless explicit consent is given, or specific conditions are met [Fog20].

4.4.1.3 Challenges

Some identified challenges to the development of DLT-based applications include interoperability, security and privacy, scalability, speed and patient engagement [KWC18].

The interoperability challenge stems from the fact that there is not yet an existing standard for developing DLT-based healthcare applications; therefore, applications developed by different vendors or on different platforms may not be able to interoperate. However interoperability is not a challenge specific to DLT per se; rather, it is a common challenge when adopting any technological innovation.

With regards to the security and privacy of DLT-based healthcare applications, there is a concern that despite the encryption techniques employed, it could still be possible to reveal the identity of a patient in a public blockchain by linking together sufficient data that are associated to that patient [RL18]. In addition, there is also the potential risk of security breaches that could arise from intentional malicious attacks to the healthcare blockchain by criminal organizations or even government agencies that could compromise the privacy of the patients. The private keys that are used for data encryption and decryption in blockchain are also prone to potential compromise which could result in unauthorized access to the stored health data.

Furthermore, there is the concern that the immutability property of blockchain does not augur well with the GDPR's "right to be forgotten," which is part of the European Union General Data Protection Regulation which stipulates that the user has the right to request for the complete erasure of the user's data [Fog20]. Since the immutability of blockchain ensures that data once saved to the blockchain cannot be deleted or altered, it could prove counterproductive when it is desirable to completely wipe out the medical history of a patient.

Scalability of blockchain-based healthcare solutions is a major challenge especially in relation to the volume of data involved. It is not optimal, or even practicable in some cases, to store the high-volume biomedical data on blockchain as this is bound to cause serious performance degradation. There is also the problem of speed as the blockchain-based processing can introduce some significant latency. For example, the validation mechanism in the current set-up of the Ethereum blockchain platform necessitates all the nodes in a network to participate in the validation process [Yli+16]. This incurs considerable processing delay, especially if the data load is significant.

One more challenge is how to engage patients in the management of their data on blockchain. Patients, especially the elderly and the young, may not be interested or able to participate in the management of their health data [RL18].

Other challenges facing the adoption of blockchain in healthcare include computational overhead and the uncertainty about who is responsible for the cost of technology implementation and who profits from it [Eng17; Udd+18]. Barriers to adopting blockchain in the health care sector include immaturity of the technology itself, insufficient skills to understand and

implement it, lack of buy-in, and lack of clear return on investment [Mam+18]. The lack of buy-in goes back to the unfamiliarity of blockchain, the negative attitudes of physicians toward the use of blockchain [Zha+18b], and the fact that not all the patients are interested in managing their health records [KWC18].

With blockchain technology, transactions are processed and verified by an automated programmable logic with predefined rules, which reduces transaction costs (ie, effort and time spent on bureaucracy) [RDR18; Zha+18a]. The complex or computation-intense systems needed in healthcare are not the best use cases for DLT [Kee+07], since performance, real-time communication, coordination, data sharing, and medical service availability are critical in life-threatening situations [KWC18].

4.4.2 Finance

4.4.2.1 Overview

Recent developments have seen the creation of digital currencies like Bitcoin, which combine new currencies with decentralized payment systems. Although the monetary aspects of digital currencies have attracted considerable attention, the distributed ledger underlying their payment systems is a significant innovation. As with money held as bank deposits, most financial assets today exist as purely digital records. This opens the possibility for DLTs to transform the financial system more generally.

In the financial sector, such as interbank payment and global financial transactions, generally a permissioned DLT is used [Yoo17]. Because of the nature of finance, reliability, stability and efficiency are priorities. Blockchains based on a permissioned DLT, where only authorized individuals can participate, are preferred. This type of DLT design has a consensus mechanism that ensures the authenticity of the transaction, so that only a small number of specific groups can participate to offset the problems of permissionless systems. First, it is to secure technological development and standardization. Permissionless DLTs are difficult to standardize because of the lack of new standard method owing to technological development, but permissioned designs are easy to agree and accept technical standards among participants. Additionally, this type can achieve efficiency and independence. Permissionless DLTs have the advantage that there is no specific power or reliance agency intervention, but the efficiency structure is lower compared to the permissioned type in consensus structure. Furthermore, in the case of the permissioned type, the transaction can be changed or modified by mutual agreement, whereas in permissionless systems it is not possible to modify the transaction recorded in the spreadsheet and can only be corrected by reverse trading. In this respect, the financial sector more often considers the adoption of permissioned DLT designs [Yoo17].

4.4.2.2 Benefits

At the beginning of the blockchain hype, financial organizations are skeptical about using the new technology. However, it is increasingly obvious how much money could be saved by processing the enormous amounts of transactions faster and more secure with less dependency on paper. Old communication processes by mail and fax for setting up syndicated loans or in the trade finance area, for example, could vanish with the help of blockchain. Interbank and

cross-border payments and settlements could take advantage of high transaction speeds with less intermediates by using cryptocurrencies.

Blockchain has the potential to be an intermediatory platform for creating trust between different parties. Counterparty verification, which is another crucial component in the banking business, could be simplified and made more secure. It will benefit from anti-money laundering algorithms implemented in smart contract logic combined with blockchain's security protocols for data protection. With regard to GDPR, data could even remain private in a customer's private blockchain "wallet" if necessary [Bet+19].

One research paper, investigating the changes in the financial sector caused by blockchains [Yoo17], suggests that, by using blockchain, major banks succeeded in the project of storing and co-management of the consumer identification information. It is a system that identifies the customer's real name, address, contact details and purpose of financial transactions so that the financial products and services provided by the financial company are not used in illegal activities such as money laundering.

Another promising effect of the introduction of the blockchain is that both the time and cost can be reduced by international transactions. If blockchains are introduced into international transactions, the payment, settlement and payment processing will be faster, which will reduce counterparties and liquidity risk. The strength of the transaction service provided by FinTech is shortened costs and time.

Furthermore, smart contract means that documents containing various contracts or information that are used offline can be safely recorded online so that contract information and information can be checked at any time and place. For example, it is possible to upload and sell information on various derivative contracts as well as personal information (e.g. car accident history and mileage, property ownership, etc.).

According to another study [Bet+19], there are generally two main perspectives, from which to look at these technologies, in context of the financial services industry - focus on the payment options or focus on smart contracts and the distributed nature of DLTs.

The second perspective focuses more on the functional aspects in a distributed ledger, namely smart contracts and the opportunity to manage personal data and access rights. Smart contracts can be used to implement business logic inside a distributed ledger and define processes as standards for all participating parties. Due to the digital nature of these implementations, smart contracts have the potential to speed up processes and achieve better adherence to contracts by algorithmically defining the intentions of processes. Ethereum is an example of a public blockchain that can carry smart contracts, which subsequently enables other blockchains to be built on top of the Ethereum blockchain with their own rules and methods. On the other hand, Ethereum itself permits payment between different parties and is therefore also part of the first perspective mentioned above.

Worldwide transactions with untrusted third parties and transactions that require perfect anonymity are the primary selling points. Blockchain's protection against forgery due to hash keys and certificate-based encryption bundled with its customizable agreement and endorsement policies to determine whether a transaction is valid makes it suitable for many applications. Another important point is the immutability of the blockchain itself, meaning it is impossible to delete or alter a transaction after it has taken place without everyone knowing it.

Other potential benefits of DLTs from a financial perspective mentioned in the literature [Fos+20] include: reduced cost and accuracy of recordkeeping, reduced fraud, more secure,

transparent and faster transactions, disintermediation, improved privacy and integrity, improved identity, space and service management, enchanced efficiency of financial services and customer trust, reduced corruption, open source, less physical infrastructure needed for the transfer of data and services.

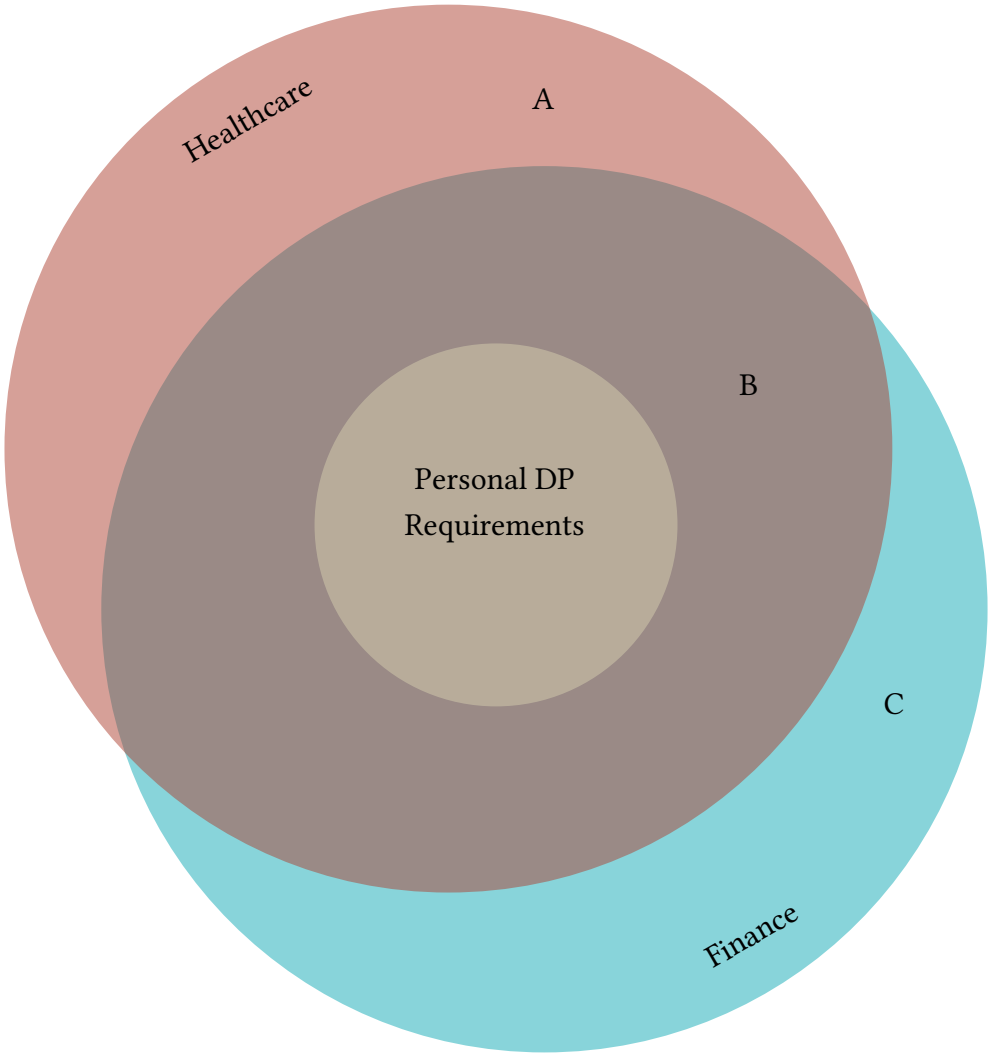
4.4.2.3 Challenges

As technology develops, consumer needs and related environments are changing. At the same time, there is an increased opportunity for individuals to be infringed by information such as hacking, and there is a strong need for blockchain technology because of the efforts of institutions trying to defend hacking. To encourage market movements, the government and related organizations should recognize the power of blockchains in individual and business transactions, public services, etc., and support them through development of original technologies and finding out best practices. However, the literature [Fos+20] suggests that many practical challenges remain in implementing blockchain solutions like, for example, limited understanding and adoption of DLTs, cost and managerial overhead, lack of government regulation, regulatory compliance issues, identity verification concerns, difficulties in illegal practices detection and tracking, scalability and throughput concerns, high computational speed and processing power demand, performance, transaction speed, network size and bandwidth concerns, high energy consumption, usability etc. Apart from the general blockchain issues, DLT approaches for personal financial information can face challenges in terms of limited cooperation between institutions and participants, rapid transformation of financial systems, reluctance to agree to standards, compliance, intelectual property, system stability, resilience and security concerns, etc.

4.4.3 Comparison

The analysis of the benefits and challenges that DLTs can bring in the domain of healthcare and finance further underline the differences we discussed in section _. Requirements such as scalability, interoperability, accessibility and compliance have nuances, and, for example, information availablity in emergency situations, consumer-institution trust and sensitivity of the data are much more important in healthcare, compared to finance. On the other hand, the attention in the domain of finance is much more focused on transaction time and cost reduction, crime and fraud prevention, as well as disintermediation.

The following diagram displays how our DP requirements relate to the answer of *Q4*:

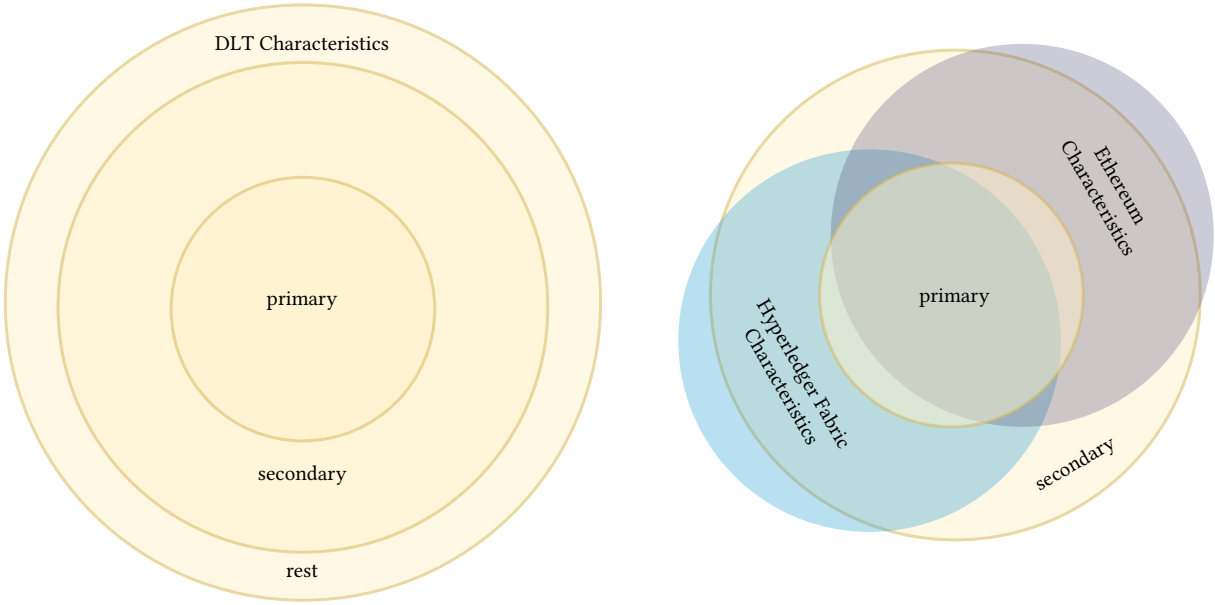


A	B	C
Emergency	Smart-contracts	Cost Reduction
Responsibility	Verifiability	Corruption Reduction
Patient Engagement	Standardisation	Crime Reduction
	Transparency	Transaction Speed
	Decentralisation	Disintermediation
	Immutability	
	Efficiency	
	Authenticity	
	Comp. Resources	
	Non-repudiation	
	Consistency	
	Stability	

4.5 DLT Characteristics

Answering *Q4* and, thereby, gaining understanding of the requirements and characteristics of DLTs, that are considered important in the available literature on healthcare and finance, has led us to the *RQ*. But firstly, we the need to take a closer look at the characteristics of DLTs and how are they defined.

As mentioned in section _, the 6 DLT properties consist of 40 different DLT characteristics. Based on the important requirements, acquired through answering *Q4*, we identified 30 out of those 40 characteristics as significant for our further investigation of DLT's suitability for DP. We further label 7 of them as "primary", due to their direct correspondence with our DP requirements and the other 23 as "secondary", which either further influence our fundamental personal DP requirements or relate to the benefits, challenges and requirements, considered important by the available literature in healthcare and finance.



Primary Characteristic	Description
User Unidentifiability	The difficulty of mapping senders and recipients in transactions to identities
Transaction Content Visibility	The ability to view the content of a transaction in a DLT design
Traceability	The extent to which transaction payloads (e.g., assets) can be traced chronologically in a DLT design
Scalability	The capability of a distributed ledger to efficiently handle decreasing or increasing amounts of required resources
Interoperability	The ability to interact between distributed ledgers and with other external data services
Compliance	The alignment of a distributed ledger and its operation with policy requirements (e.g., regulations or industry standards)
Ease of Use	The simplicity of accessing and working with a distributed ledger

Secondary Characteristic	Description
Node Controller Verification	The extent to which the identity of validating node controllers is verified prior to joining a distributed ledger
Resource Consumption	The computational efforts required to operate a distributed ledger (e.g., for transaction validation, block creation, or storing the distributed ledger)
Throughput	The maximum number of transactions that can be appended to a distributed ledger in a given time interval
Maintainability	The degree of effectiveness and efficiency with which a distributed ledger can be kept operational
Token Support	The possible uses of tokens within a distributed ledger (e.g., security token, stable coin, or utility token)
Turing-complete Smart Contracts	The support of Turing-complete smart contracts within a DLT design
Confidentiality	The degree to which unauthorized access to data is prevented
Integrity	The degree to which transactions in the distributed ledger are protected against unauthorized (or unintended) modification or deletion
Availability	The probability that a distributed ledger is operating correctly at any point in time

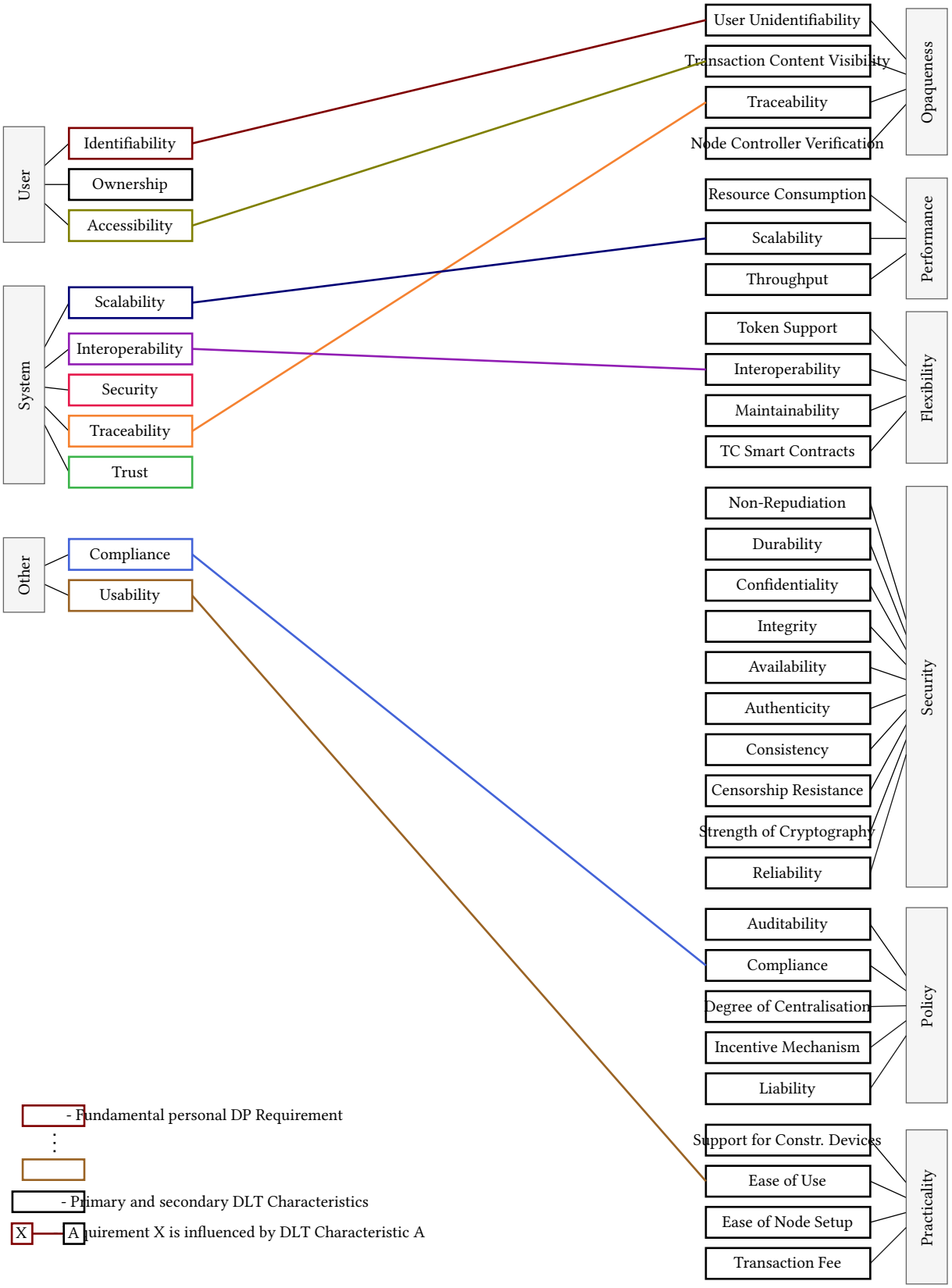
Non-Repudiation	The difficulty of denying participation in transactions
Durability	The property that data committed to the distributed ledger will not be lost
Authenticity	The degree to which the correctness of data that is stored on a distributed ledger can be verified
Consistency	The absence of contradictions across the states of the ledger maintained by all nodes participating in the distributed ledger
Censorship Resistance	The probability that a transaction in a distributed ledger will be intentionally aborted by a third party or processed with malicious modifications
Reliability	The ability of a system or component to perform its required functions under stated conditions for a specified time
Strength of Cryptography	The difficulty of breaking the cryptographic algorithms used in the DLT design
Degree of Decentralization	The number of independent validating node controllers reduced by the number of controllers that control more than average validating nodes divided by the total number of node controllers in the network.
Incentive Mechanism	A structure in place to motivate node behavior that ensures viable long-term operation of a distributed ledger (e.g., by contributing computational resources)
Liability	The existence of a natural or legal person that can be subjected to litigation with respect to the distributed ledger
Auditability	The degree to which an independent third party (e.g., state institution, certification authority) can assess the functionality of a distributed ledger
Support for Constrained Devices	The extent to which devices with limited computing capacities (e.g., sensor beacons) can participate in a distributed ledger
Transaction Fee	The price transaction initiators can or must pay for the processing of transactions
Ease of Node Setup	The ease of configuring and adding a new (or previously crashed) node to the distributed ledger

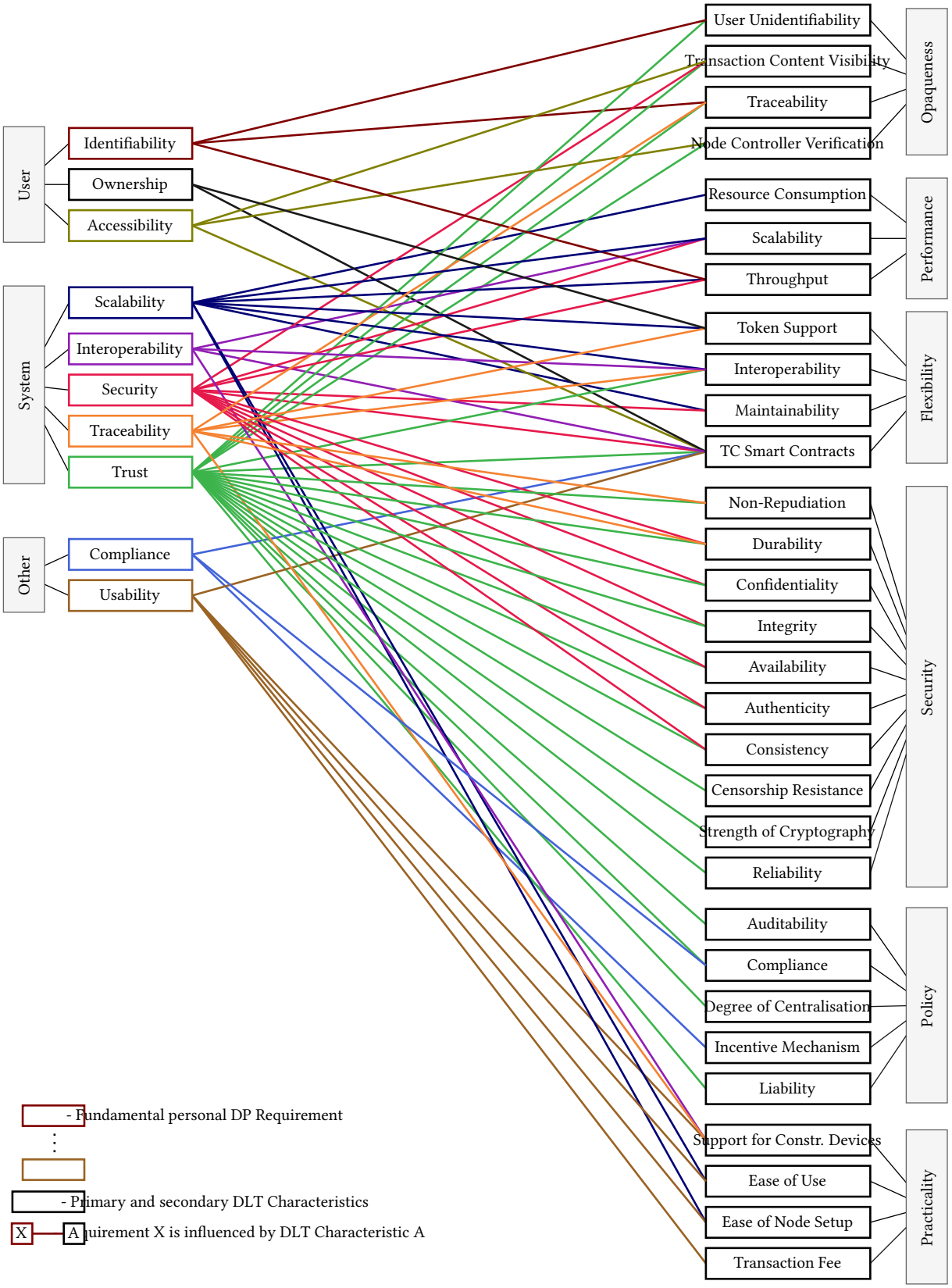
The 7 "primary" DLT characteristics (fig) correspond directly to some of our defined DP requirements. The other 23 "secondary" DLT characteristics (fig) that we identified are rather concrete and specific. Furthermore, our defined DP requirements (table) can be influenced by one or more DLT characteristics. In the following section of our work we aim to present a mapping and display which DLT characteristics seem to be related to which DP requirements.

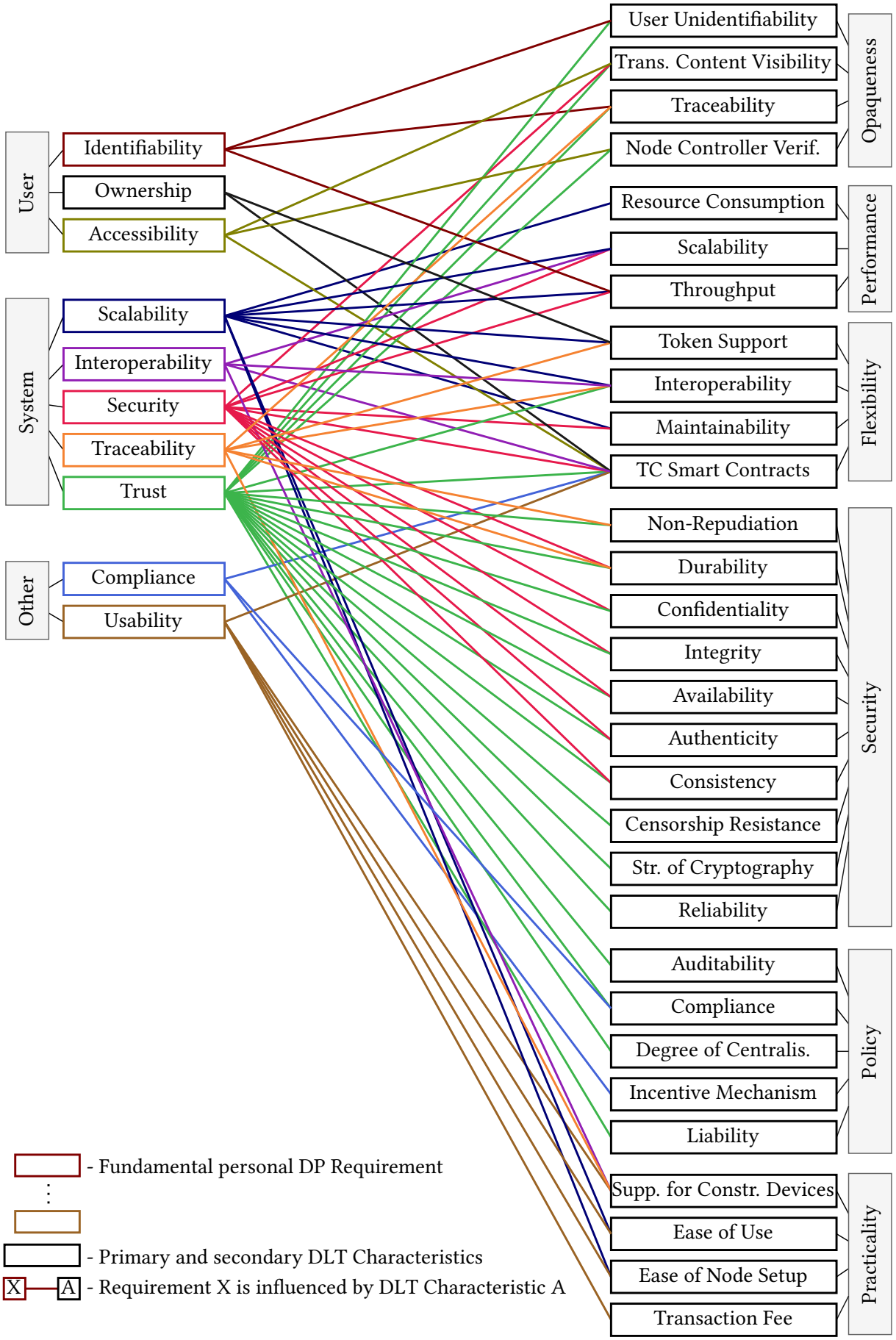
4.6 Mapping

RQ: What are the characteristics of DLTs that make them suitable for personal data provenance in healthcare and finance?

The DLTs' characteristics that we labeled as "primary" can be easily mapped to our initial DP requirements, due to their similar definitions and concerns (fig prim). From the mapping we can conclude that most of our DP requirements can be addressed, influenced by or are generally considered important by DLT approaches and directly correspond to some characteristics. However, by analysing the primary relationships together with the secondary, we can gain a better understanding of both which DP requirements DLTs 'prefer' to fulfill and which DLT characteristics are of bigger importance for DP (fig all).


User
Identifiability
Ownership
Accessibility
System
Scalability
Interoperability
Security
Traceability
Trust
Other
Compliance
Usability
User Unidentifiability
Transaction Content Visibility
Traceability
Node Controller Verification
Opaqueness
Resource Consumption
Scalability
Throughput
Performance
Token Support
Interoperability
Maintainability
TC Smart Contracts
Flexibility
Non-Repudiation
Durability
Confidentiality
Integrity
Availability
Authenticity
Consistency
Censorship Resistance
Strength of Cryptography
Reliability
Security
Auditability
Compliance
Degree of Centralisation
Incentive Mechanism
Liability
Policy
Support for Constr. Devices
Ease of Use
Ease of Node Setup
Transaction Fee
Practicality
- Fundamental personal DP Requirement
- Primary and secondary DLT Characteristics
X
A
uirement X is influenced by DLT Characteristic A


User
System
Other
Identifiability
Ownership
Accessibility
Scalability
Interoperability
Security
Traceability
Trust
Compliance
Usability
User Unidentifiability
Transaction Content Visibility
Traceability
Node Controller Verification
Resource Consumption
Scalability
Throughput
Token Support
Interoperability
Maintainability
TC Smart Contracts
Non-Repudiation
Durability
Confidentiality
Integrity
Availability
Authenticity
Consistency
Censorship Resistance
Strength of Cryptography
Reliability
Auditability
Compliance
Degree of Centralisation
Incentive Mechanism
Liability
Support for Constr. Devices
Ease of Use
Ease of Node Setup
Transaction Fee
Opaqueness
Performance
Flexibility
Security
Policy
Practicality
- Fundamental personal DP Requirement
- Primary and secondary DLT Characteristics
X
A
uirement X is influenced by DLT Characteristic A


User
Identifiability
Ownership
Accessibility
System
Scalability
Interoperability
Security
Traceability
Trust
Other
Compliance
Usability
User Unidentifiability
Trans. Content Visibility
Traceability
Node Controller Verif.
Opaqueness
Resource Consumption
Scalability
Throughput
Performance
Token Support
Interoperability
Maintainability
TC Smart Contracts
Flexibility
Non-Repudiation
Durability
Confidentiality
Integrity
Availability
Authenticity
Consistency
Censorship Resistance
Str. of Cryptography
Reliability
Security
Auditability
Compliance
Degree of Centralis.
Incentive Mechanism
Liability
Policy
Supp. for Constr. Devices
Ease of Use
Ease of Node Setup
Transaction Fee
Practicality
- Fundamental personal DP Requirement
- Primary and secondary DLT Characteristics
X A - Requirement X is influenced by DLT Characteristic A

4.6.1 Influenced Requirements

This mapping shows us that DLTs can potentially satisfy all of our DP requirements to some degree. On the first look we can't help but notice that the **Usability** and System level DP requirements **Scalability**, **Traceability** and especially **Security** and **Trust** are strongly influenced by a great number of DLT characteristics.

Scalability, as mentioned in previoius sections, is a requirement of great importance in both use cases, discussed in this paper. It is influenced by a number of characteristics: the computational efforts required to operate a distributed ledger (Resource Consumption); the maximum number of transactions that can be appended in a given time interval (Throughput); the degree of effectiveness and efficiency with which the network can be kept operational; the ability of the technology to interact with other DLTs and external data services (Interoperability); the ease of working with and configuring or adding new nodes to the network (Ease of Use, Ease of Node Setup); as well as having a token to support various services and features (Token Support).

Interoperability, as previously mentioined, remains one of the biggest challenges. It is influenced by the extent to which devices with limited computing capacities can participate in the network (Support of Constrained Devices) and more importantly, by the DLT design's ability to support smart contracts (Turing-complete Smart Contracts).

Traceability and provenance are often used synonymously. The possible use of tokens within the DLT (Token Support), the DLT's ability to interact with other systems (Interoperability), the difficulty of denying participation in transactions (Non-repudiation) and the support of devices with limited computing capacities (Support of Constrained Devices) positively influence the ability to trace information in the system.

Security is one of the main advantages that DLTs propose. This is the second most influenced DP requirement and is related to 11 DLT characteristics, namely: the ability to view the content of a transaction in a DLT design (Transaction Content Visibility); the property that data committed to the distributed ledger will not be lost (Durability); the degree to which unauthorized access to data is prevented (Confidentiality); The degree to which transactions in the distributed ledger are protected against unauthorized modification or deletion (Integrity); the probability that a distributed ledger is operating correctly at any point in time (Availability); the degree to which the correctness of data that is stored on a distributed ledger can be verified (Authenticity); the absence of contradictions across the states of the ledger maintained by all nodes participating in the distributed ledger (Consistency); as well as the previously mentioned characteristics such as Scalability, Throughput, Maintainability and Turing-complete Smart Contracts.

What DLTs can bring in DP is **Trust**. Out of the 6 properties of DLTs, 3 of them have an impact on consumer's Trust in the system - opaqueness, security and policy (3 properties, 19 characteristics). Opaqueness includes characteristics such as the difficulty of mapping senders and recipients in transactions to identities (User Unidentifiability); the extent to which transaction payloads can be traced chronologically in a DLT design (Traceability); the extent to which the identity of validating node controllers is verified prior to joining a distributed ledger (Node Controller Verification); as well as Transaction Content Visibility. The Security property includes characteristics such as the probability that a transaction in a distributed ledger will be intentionally aborted by a third party or processed with malicious modifications

(Censorship Resistance); the difficulty of breaking the cryptographic algorithms used in the DLT design (Strenght of Cryptography); the ability of a system or component to perform its required functions under stated conditions for a specified time (Reliability); as well as the previously mentioned characteristics like Confidentiality, Integrity, Availability, Authenticity, Consistency, Dirability and Non-repudiation. The Policy property consists of characteristics such as degree to which an independent third party (e.g., state institution, certification authority) can assess the functionality of a distributed ledger (Auditability); the alignment of a distributed ledger and its operation with policy requirements (e.g., regulations or industry standards) (Compliance); the number of independent validating node controllers reduced by the number of controllers that control more than average validating nodes divided by the total number of node controllers in the DLT network (Degree of Decentralisation); the existence of a structure in place to motivate node behavior that ensures viable long-term operation of a distributed ledger (Incentive Mechanism) and the existence of a natural or legal person that can be subjected to litigation with respect to the distributed ledger (Liability).

Additionally, 5 of the DLT characteristics have an impact on the **Usability** requirement. These are Turing-Complete Smart Contracts, Support for Constrained Devices, Ease of use, Ease of Node Setup and the price that transaction initiatiors can or must pay for the processing of transcations (Transaction Fee).

There remain 5 DP requirements, which are not influenced by as many DLT characteristics as the other half that we discussed. **Identifiability**, for example, is impacted by DLT's characteristics of User Unidentifiability, Traceability and Throughput. **Accessibility** is influenced by Transaction Content Visibility, Node Controller Verification and the DLT's ability to support Turing-complete Smart Contracts. **Ownership** can be influenced by the possible uses of tokens within the network (e.g., security token, stable coin, or utility token) and Smart Contracts and **Compliance** is additionally influenced by the alignment of a distributed ledger and its operation with policy requirements and the existence of astructure in place to motivate node behavior that ensures viable long-term operation of a distributed ledger (Compliance and Incentive Mechanism).

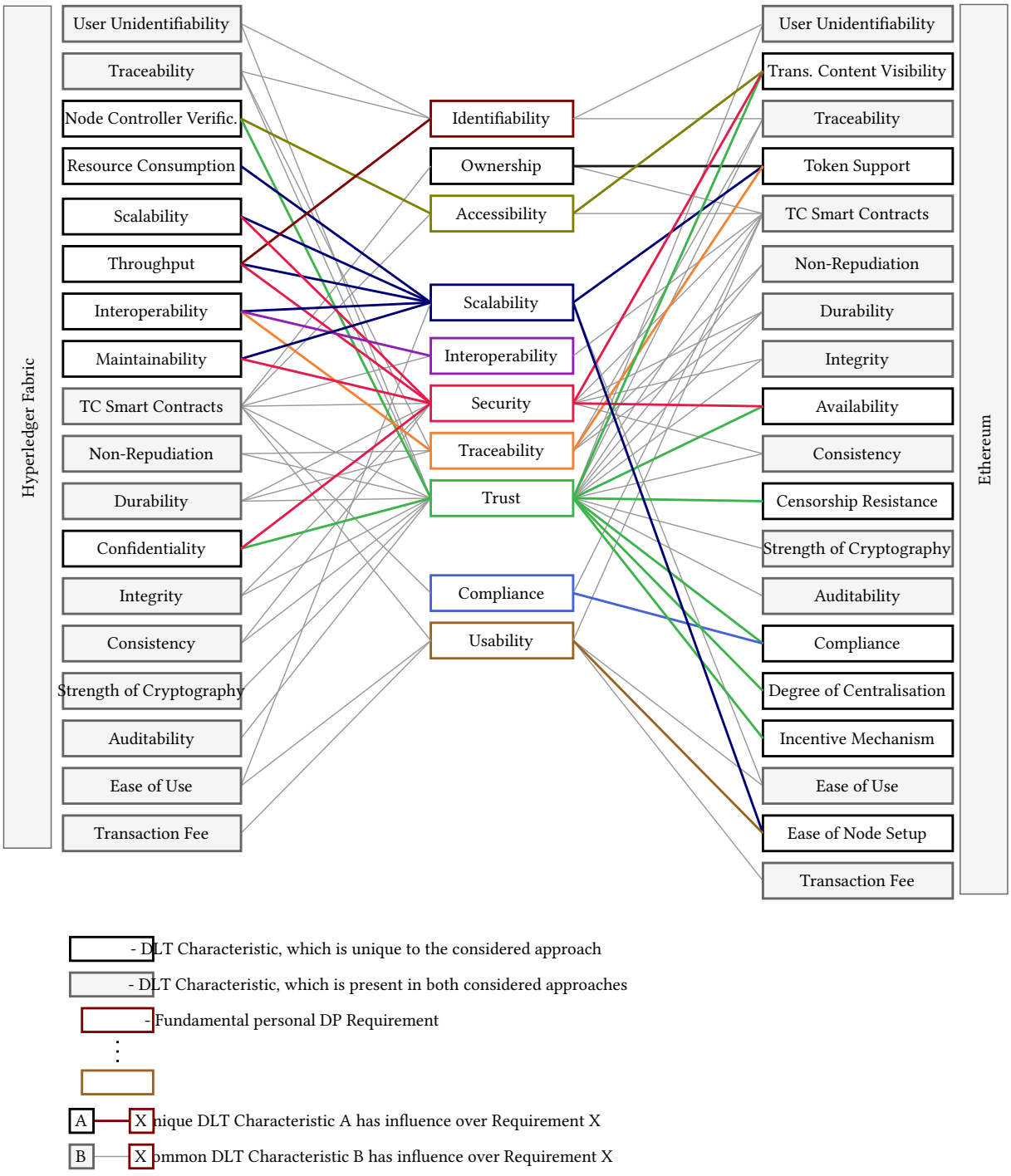
subsectionInfluential Characteristics

From the mapping (fig), we can observe that DLTs' ability to support Turing-complete Smart Contracts is the most important DLT characteristic for a successful DP approach, followed by DLTs' degree of Interoperability, Scalability, Traceability, Throughput, etc. Rightfully so, smart contracts can be used to satisfy numerous requirements, which the considered approaches in our work don't claim have as underlying features, such as **Accessibility**, **Ownership**, **Compliance** and higher **Traceability** and **Usability**.

The analysis of the mapping provides and overview as to which DLT characteristics are beneficial or suitable for personal DP in general. However, this work considers two specific use cases (healthcare and finance), which have their own requirements, and two approaches (Hyperledger Fabric and Ethereum), which do not exhibit all of the DLT characteristics that are considered benefitial for personal DP.

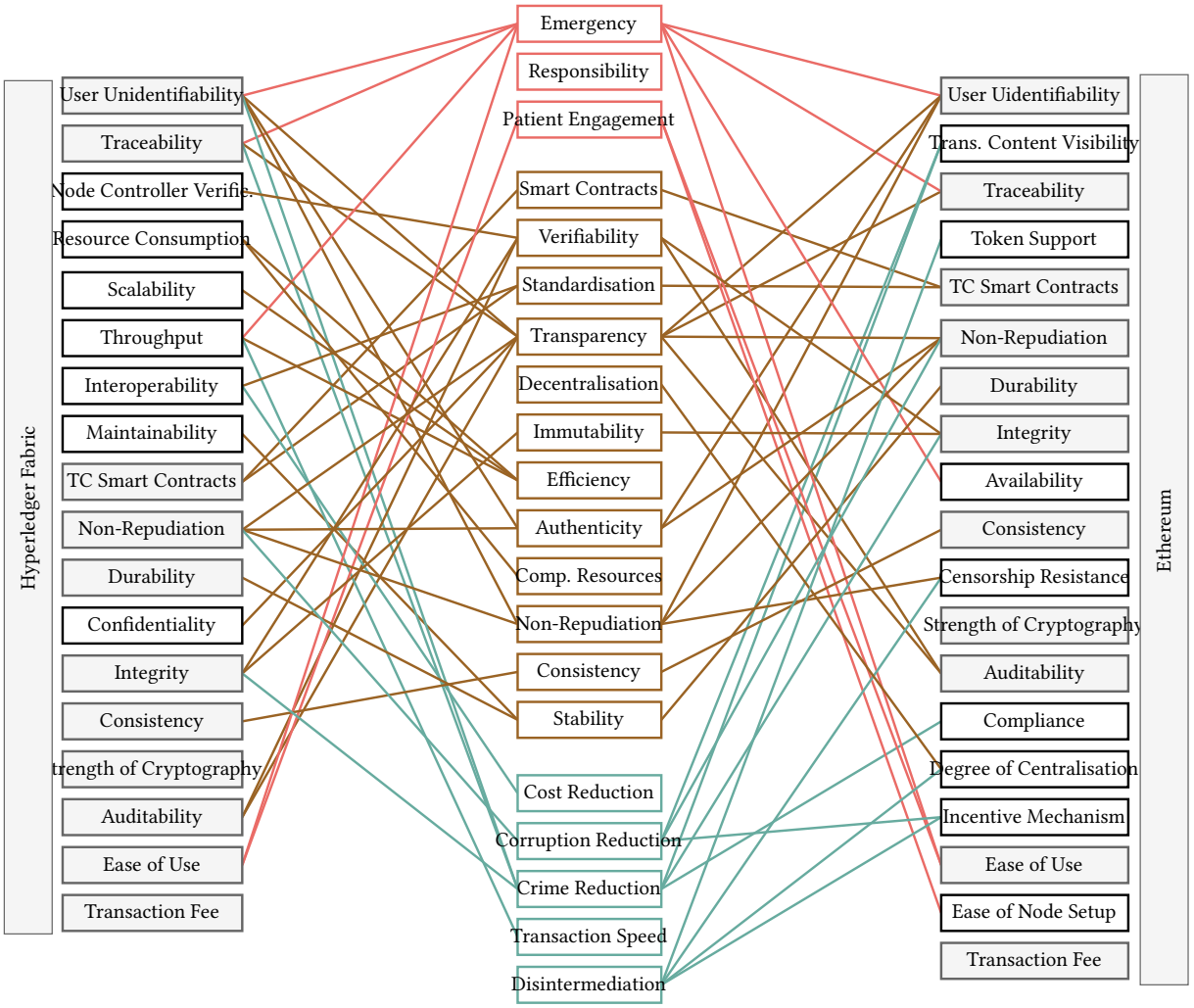
4.6.2 CA Characteristics to DP Requirements

After analysing Hyperledger Fabric's and Ethereum's whitepapers, other available literature on the appliaction of these approaches in healthcare and finance, as well as consulting with individuals with thorough knowledge in these approaches, we concluded that out of the 30 previously selecteed DLT characteristics, the considered approaches exhibit 18 and 19 characteristics, respectively, 11 of which they share. The characteristics, that are unique to each approach, we mapped to the DP requirements with a corresponding color, in order to gain an understanding in which areas the DLT designs excel and what benefits and trade-offs they can bring against each other. What we observed is that, similarly to our previous analysis, Fabric can provide potentially better Identifiability, Scalability, Interoperability and Security, whereas Ethereum brings greater benefits in terms of Trust and Compliance. However, it is important to note that the right side characteristics describe the public version of Ethereum, as the private approach is quite similar to Fabric. Additionally, other features of the considered approaches like, for example, smart contract upgradability or side-chains and sharding are not represented through these characteristics.



4.6.3 CA Characteristics to UC Requirements

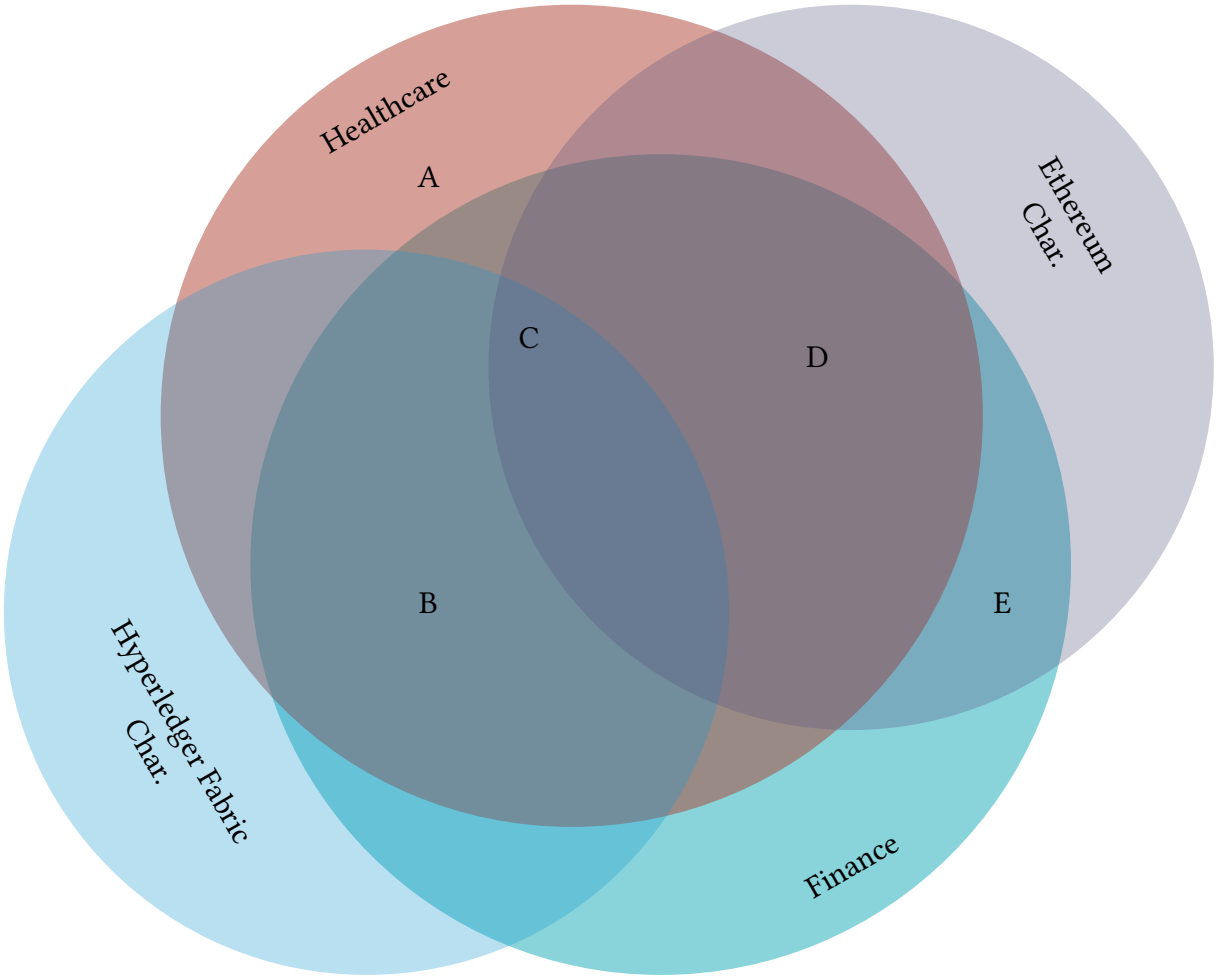
The considered approaches in this work seem to also have an impact on the majority of the requirements that we acquired through answering *Q4*. All of the common use case requirements are influenced by either Fabric, Ethereum or both. It is important, however, that there seems to be no characteristic that can impact the Responsibility requirement in healthcare (map).



- DLT Characteristic, which is unique to the considered approach
- DLT Characteristic, which is present in both considered approaches
- Requirement, which is unique to the healthcare use case
- Requirement, which is unique to the financial use case
- Requirement, which is present in both use cases

A – Unique DLT Characteristic A has influence over Requirement X

B – Common DLT Characteristic A has influence over Requirement X



A	B	C	D	E
Responsibility	Efficiency	Smart Contracts	Decentralisation	Disintermediation
	Comp. Resources	Verifiability		
	Cost Reduction	Standardisation		
	Trans. Speed	Transparency		
		Immutability		
		Authenticity		
		Non-repudiation		
		Consistency		
		Stability		

5 Discussion

5.1 Principle Findings

This research reveals 26 DLT characteristics, that make DLTs suitable for personal DP, based on mappings of 2 DLT approaches' characteristics (Hyperledger Fabric and Ethereum) to 10 identified fundamental personal DP requirements (table) and another 22 additional requirements (). The fundamental DP requirements we analysed through the lens of 2 different use cases (healthcare and finance); the additional requirements, we determined by analysing which features and characteristics were considered beneficial, challenging or generally important by the available literature on DLT-based DP in both use cases.

Among the fundamental DP requirements, Identifiability, Accessibility, Scalability, Interoperability, Traceability, Compliance and Usability seem to be directly corresponding to specific DLT characteristics. The most influenced requirements are Trust and Security (19 and 11 relations to DLT characteristics), followed by Scalability, Traceability and Interoperability, signifying DLTs' suitability for DP in terms of system level DP requirements. In terms of characteristics, Turing-Complete Smart Contracts is the most influencial DLT characteristic. Out of the considered approaches, Hyperledger Fabric seems to provide better Identifiability, Scalability (in terms of data), Interoperability and Security (in terms of confidentiality), whereas Ethereum seems to bring more Trust, Compliance and Usability.

In terms of the additional 22 requirements extracted from the literature, Fabric doesn't seem to satisfy the Decentralisation and Disintermediation requirement, Ethereum fails to meet the Cost Reduction, Transaction speed and Efficiency requirements, whereas both considered approaches don't seem to address the Responsibility requirement in healthcare. In this work, one approach doesn't seem to be overall better than the other. The considered approaches seem to benefit from different characteristics, which leads to trade-offs, that should be considered when dealing with DP in different use cases such as healthcare and finance.

5.2 Implications for Practice

Our research can assist practitioners in obtaining insights into the viability of Hyperledger Fabric, Ethereum or DLT designs, exhibiting similar characteristics, for applications and possible impact of DLTs in terms of personal DP in the domain of healthcare and finance. This work can aid the decision making process of selecting DLT designs for application use cases, similar to ours, under the consideration of personal DP requirements and DLT characteristics.

Our mappings reveal trade-offs, which are useful to be aware of potential benefits and drawbacks before developing an application in the discussed domains. Such assessments eventually facilitate avoidance of unsuitable DLT designs or unsuitable use cases and consequent waste of resources. "Careful DLT design selection and application development becomes crucial

to ensure that DLT's unique advantages can be achieved, ultimately pushing DLT from a hype to a critical information infrastructure [DLS19] for future businesses and societies" [Kan+20].

5.3 Implications for Research

Our analysis of Hyperledger Fabric and Ethereum characteristics' suitability for personal DP in the domain of healthcare and finance can be used to develop a better understanding of important aspects of the relationship between DLTs and personal DP.

The review, mappings and classifications proposed in this study offer useful insights into the research on Ethereum, Hyperledger Fabric, and personal DP in the field of healthcare and finance. They place all four concepts in one paper, making a comparative analysis easier for readers. In addition, the proposed definitions and findings can be used as a research agenda in Ethereum, Hyperledger Fabric, healthcare and finance orientations and related discussions, amid the perception that further research in this area should be aligned to their rapid development.

This work can serve as motivation for research on ways and means of overcoming the explored drawbacks and challenges of the considered approaches. Another angle of research may consist of investigating suitable DLT design beyond Fabric and Ethereum or personal DP use cases beyond healthcare and finance.

5.4 Limitations and Future Work

Nevertheless, our study comes with limitations. Personal DP requirements and DLT characteristics were identified in a literature review in the field of DLT, DP healthcare and finance. Analyzed DLT and DP concepts are limited to already published scientific articles and mainly focused on blockchain and DP of personal information. We limit our overview of personal DP, as well as use case specific requirements and DLT characteristics to those of particular interest in extant research on DP and DLT in the domain of healthcare and finance. The DLT characteristics and related trade-offs, as well as the DP requirements are also corroborated by multiple whitepapers of DLT designs such as Bitcoin [DL], Ethereum [], or Hyperledger Fabric [DMH17] and other surveys, reviews, taxonomies, comparisons, case studies and reports. Most of the analyzed research articles developed applications on Ethereum or Hyperledger Fabric, which makes this work (Fabric and Ethereum characteristics' suitability for personal DP in healthcare and finance) only partially generalizable to other DLT designs and use cases. Also, the requirements discussed, are based on rather simple definitions from the literature, therefore, more in-depth analysis of each individual requirements is necessary, in order to extract more valuable comparisons, suppositions and conclusions. Additionaly, when we considered that a requirement is influenced by a characteristic, that doesn't necessarily mean that the requirement is satisfied and it is not discussed to what degree it is impacted. What exact degree of influence or impact does a characteristic have on a requirement and personal DP in general, needs to be further investigated.

While we analyzed relationships between DLT characteristics and DP requirements, we predominantly focused on synergistic and potentially positive influences. We acknowledge

that the relationships developed through the mappings might also lead to negative effects. The resulting trade-offs are also between approaches that represent only 2 out of 4 types of DLT desins.

DLTs' application in healthcare is still an emerging field, compared to finance. There is need for researchers to develop more prototypes and proof-of-concepts to deepen the understanding and maturity of the technology in relation to its application in healthcare. Many of the analysed frameworks, concepts, models and architectures need to be implemented and tested to evaluate their strengths and weaknesses.

Interoperability, scalability and usability remain some of the biggest challenges and turn out as an important avenue for future research in the field of DLT and DP, in order to overcome the standardisation, emergency, engagement, as well as cost reduction problems in healthcare and finance. Interoperable DLTs can increase flexibility and help laverage the benefits of different DLT designs while avoiding their drawbacks.

6 Conclusion

In this research we determined which DLT characteristics are suitable for personal DP through mapping the characteristics of Hyperledger Fabric and Ethereum to fundamental personal DP requirements, as well as specific use case requirements in the field of healthcare and finance. While both use cases share the fundamental, as well as other secondary DP requirements, healthcare seems to be focused more on patient engagement, resposibility and availability in emeregency situations, whereas finance places higher importance on disintermediation, transaction speed, cost, corruption and crime reduction.

We concluded that neither of our considered approaches is necessarily better than the other and they both exhibit characteristics which are beneficial in different situations. Turing-complete Smart Contracts turned out to be the most influencial DLT characteristic, which both considered approaches have. Apart from that, Hyperledger Fabric seems to provide better Identifiability, Scalability (in terms of transactions per second), Interoperability and Security (in terms of confidentiality), whereas Ethereum seems to bring more Trust, Compliance and Usability. There is, however, a trend in the literature that we analysed, which shows that researchers prefer permissioned DLT designs, which can be largely motivated by the improvements in throughput, scalability, performance and enchanced confidentiality due to the restricted access to the ledger. However, the need for interoperability remains as one of the biggest challenges in order to overcome the standardisation, emergency, engagement, as well as cost reduction problems in the field of healthcare and finance.

Bibliography

[] *7 Data Privacy Trends for 2021 – Data Privacy Manager*. URL: https://dataprivacymanager.net/7-data-privacy-trends-for-2020/ (visited on 2021-06-05).

[] *Chapter 3 (Art. 12-23) Archives*. en-US. URL: https://gdpr.eu/tag/chapter-3/ (visited on 2021-06-05).

[] *Ethereum Whitepaper*. en. URL: https://ethereum.org (visited on 2021-11-15).

[] *Getting my personal data out of Facebook*. en. URL: https://ruben.verborgh.org/facebook/ (visited on 2021-06-05).

[] *Healthcare data volume globally 2020 forecast*. en. URL: https://www.statista.com/statistics/1037970/global-healthcare-data-volume/ (visited on 2021-12-04).

[] *Matomo Analytics - The Google Analytics alternative that protects your data*. URL: https://matomo.org/ (visited on 2021-06-05).

[] *Number of fintech startups globally by region 2021*. en. URL: https://www.statista.com/statistics/893954/number-fintech-startups-by-region/ (visited on 2021-12-04).

[] *Plasma - EthHub*. URL: https://docs.ethhub.io/ethereum-roadmap/layer-2-scaling/plasma/ (visited on 2021-12-03).

[] *SecondProvenanceChallenge < Challenge < TWiki*. URL: https://openprovenance.org/provenance-challenge/SecondProvenanceChallenge.html (visited on 2021-11-13).

[] *Sharding FAQs · ethereum/wiki Wiki*. en. URL: https://github.com/ethereum/wiki (visited on 2021-12-03).

[] *Use Case Anonymous Information - XG Provenance Wiki*. URL: https://www.w3.org/2005/Incubator/prov/wiki/Use_Case_Anonymous_Information (visited on 2021-11-14).

[] *Use Case private data use - XG Provenance Wiki*. URL: https://www.w3.org/2005/Incubator/prov/wiki/Use_Case_private_data_use (visited on 2021-11-14).

[] *Walmart Case Study*. en-US. URL: https://www.hyperledger.org/learn/publications/walmart-case-study (visited on 2021-06-05).

[17] "IoT Data Provenance Implementation Challenges". en. In: *Procedia Computer Science* 109 (Jan. 2017). Publisher: Elsevier, pp. 1134–1139. ISSN: 1877-0509. DOI: 10.1016/j.procs.2017.05.436. URL: https://www.sciencedirect.com/science/article/pii/S1877050917311183 (visited on 2021-06-05).

[18] *California Consumer Privacy Act (CCPA)*. en. Oct. 2018. URL: https://oag.ca.gov/privacy/ccpa (visited on 2021-06-06).

[19] *What is the LGPD? Brazil's version of the GDPR*. en-US. Section: News & Updates. July 2019. URL: https://gdpr.eu/gdpr-vs-lgpd/ (visited on 2021-06-06).

[21a] *Ethereum Proof-of-Stake Consensus Specifications*. original-date: 2018-09-20T05:12:54Z. Dec. 2021. URL: https://github.com/ethereum/consensus-specs (visited on 2021-12-03).

[21b] *Polkadot Paper*. original-date: 2017-07-24T11:37:49Z. June 2021. URL: https://github.com/w3f/polkadot-white-paper/blob/161787ea0e01aef43b040d9737915218bf19f75/PolkaDotPaper.pdf (visited on 2021-12-03).

[Adi+17] Mohammad Adibuzzaman et al. "Big data in healthcare–the promises, challenges and opportunities from a research perspective: A case study with a model database". In: *AMIA Annual Symposium Proceedings*. Vol. 2017. American Medical Informatics Association. 2017, p. 384.

[AG05] Alessandro Acquisti and Jens Grossklags. "Privacy and rationality in individual decision making". In: *IEEE security & privacy* 3.1 (2005), pp. 26–33.

[AHS02] Helen Allen, John Hawkins, and Setsuya Sato. "Electronic trading and its implications for financial systems". In: *Technology and Finance*. Routledge, 2002, pp. 213–247.

[AI19] Nizamuddin Ariffin and Ahmad Zuhairi Ismail. "The design and implementation of trade finance application based on hyperledger fabric permissioned blockchain platform". In: *2019 International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. IEEE. 2019, pp. 488–493.

[AME19] Cornelius C Agbo, Qusay H Mahmoud, and J Mikael Eklund. "Blockchain technology in healthcare: a systematic review". In: *Healthcare*. Vol. 7. 2. Multidisciplinary Digital Publishing Institute. 2019, p. 56.

[AMS02] Franklin Allen, James McAndrews, and Philip Strahan. "E-Finance: An Introduction". In: *Journal of Financial Services Research* 22.1 (Aug. 2002), pp. 5–27. ISSN: 1573-0735. DOI: 10.1023/A:1016007126394. URL: https://doi.org/10.1023/A:1016007126394.

[Asg+12] Muhammad Rizwan Asghar et al. "Securing Data Provenance in the Cloud". In: *Open Problems in Network Security*. Ed. by Jan Camenisch and Dogan Kesdogan. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 145–160. ISBN: 978-3-642-27585-2.

[ASS17] Thibaud Antignac, David Sands, and Gerardo Schneider. "Data minimisation: a language-based approach". In: *IFIP International Conference on ICT Systems Security and Privacy Protection*. Springer. 2017, pp. 442–456.

[AW21] MA Hannan Bin Azhar and Robert Vause Whitehead. "A study of user experiences and network analysis on anonymity and traceability of bitcoin transactions". In: *EAI Endorsed Transactions on Security and Safety* (2021).

[Bal17] Arati Baliga. "Understanding blockchain consensus models". In: *Persistent* 4 (2017), pp. 1–14.

[BC14] Sebastian K Boell and Dubravka Cecez-Kecmanovic. "A hermeneutic approach for conducting literature reviews and literature searches". In: *Communications of the Association for information Systems* 34.1 (2014), p. 12.

[BC16] Dalel Bouslimi and Gouenou Coatrieux. "A crypto-watermarking system for ensuring reliability control and traceability of medical images". In: *Signal Processing: Image Communication* 47 (2016), pp. 160–169.

[Ber14] Teklit Hailemichael Berhe. "Conflict between anti-money laundering and anti-terrorism finance laws requirements and bank secrecy and confidentiality laws". PhD thesis. Institute of Advanced Legal Studies, 2014.

[Ber17] Jonatan Bergquist. *Blockchain Technology and Smart Contracts: Privacy-Preserving Tools*. 2017.

[Bet+19] Martina Bettio et al. "Hyperledger fabric as a blockchain framework in the financial industry". In: *The Impact of Digital Transformation and FinTech on the Finance Professional*. Springer, 2019, pp. 29–44.

[BFV19] Mirko Bez, Giacomo Fornari, and Tullio Vardanega. "The scalability challenge of ethereum: An initial quantitative analysis". In: *2019 ieee international conference on service-oriented system engineering (sose)*. IEEE. 2019, pp. 167–176.

[BGS05] Bettina Berendt, Oliver Günther, and Sarah Spiekermann. "Privacy in e-commerce: Stated preferences vs. actual behavior". In: *Communications of the ACM* 48.4 (2005), pp. 101–106.

[BHS02] France Belanger, Janine S Hiller, and Wanda J Smith. "Trustworthiness in electronic commerce: the role of privacy, security, and site attributes". In: *The journal of strategic Information Systems* 11.3-4 (2002), pp. 245–270.

[BKB16] Christoph Bier, Kay Kühne, and Jürgen Beyerer. "PrivacyInsight: the next generation privacy dashboard". In: *Annual Privacy Forum*. Springer. 2016, pp. 135–152.

[BM16] Jürgen Bott and Udo Milkau. "Towards a framework for the evaluation and design of distributed ledger technologies in banking and payments". In: *Journal of Payments Strategy & Systems* 10.2 (2016), pp. 153–171.

[Boa11] John R Boatright. "Trust and integrity in banking". In: *Ethical perspectives* 18.4 (2011), p. 473.

[Boi+21] Frederic Boissay et al. "Big techs in finance: on the new nexus between data privacy and competition". In: *The Palgrave Handbook of Technological Finance*. Springer, 2021, pp. 855–875.

[Bos01] Biagio Bossone. "Do banks have a future?: A study on banking and finance as we move into the third millennium". In: *Journal of banking & finance* 25.12 (2001), pp. 2239–2276.

[Bra05] David J Brailer. "Interoperability: The Key To The Future Health Care System: Interoperability will bind together a wide network of real-time, life-critical data that not only transform but become health care." In: *Health affairs* 24.Suppl1 (2005), W5–19.

[BS08] Mary E Barth and Katherine Schipper. "Financial reporting transparency". In: *Journal of Accounting, Auditing & Finance* 23.2 (2008), pp. 173–190.

[BT19] Peter Buneman and Wang-Chiew Tan. "Data Provenance: What next?" In: *ACM SIGMOD Record* 47.3 (Feb. 2019), pp. 5–16. ISSN: 0163-5808. DOI: 10.1145/3316416.3316418. URL: https://doi.org/10.1145/3316416.3316418 (visited on 2021-06-06).

[But20] Tom Butler. "What's Next in the Digital Transformation of Financial Industry?" In: *IT Professional* 22.1 (2020), pp. 29–33.

[CAG02] Lorrie Faith Cranor, Manjula Arjula, and Praveen Guduru. "Use of a P3P user agent by early adopters". In: *Proceedings of the 2002 ACM workshop on Privacy in the Electronic Society*. 2002, pp. 1–10.

[CFG07a] Luis V Casaló, Carlos Flavián, and Miguel Guinalıu. "The role of security, privacy, usability and reputation in the development of online banking". In: *Online Information Review* (2007).

[CFG07b] Luis V Casaló, Carlos Flavián, and Miguel Guinalıu. "The role of security, privacy, usability and reputation in the development of online banking". In: *Online Information Review* (2007).

[CFG08] Luis V Casaló, Carlos Flavián, and Miguel Guinalıu. "The role of satisfaction and website usability in developing customer loyalty and positive word-of-mouth in the e-banking services". In: *International journal of bank marketing* (2008).

[Cha85] David Chaum. "Security without identification: Transaction systems to make big brother obsolete". In: *Communications of the ACM* 28.10 (1985), pp. 1030–1044.

[Cho+19] Mohammad Jabed Morshed Chowdhury et al. "A comparative analysis of distributed ledger technology platforms". In: *IEEE Access* 7 (2019), pp. 167930–167943.

[CL+99] Miguel Castro, Barbara Liskov, et al. "Practical byzantine fault tolerance". In: *OSDI*. Vol. 99. 1999. 1999, pp. 173–186.

[Cle19] Clearmatics. *Ion Stage 2: Toward a General Interoperability Protocol (Part 1)*. en. Feb. 2019. URL: https://medium.com/clearmatics/ion-stage-2-toward-a-general-interoperability-protocol-part-1-d12b9d7316d3 (visited on 2021-12-03).

[Cow02] Christopher J Cowton. "Integrity, responsibility and affinity: three aspects of ethics in banking". In: *Business Ethics: A European Review* 11.4 (2002), pp. 393–400.

[Cro+16a] Kyle Croman et al. "On scaling decentralized blockchains". In: *International conference on financial cryptography and data security*. Springer. 2016, pp. 106–125.

[Cro+16b] Michael Crosby et al. "Blockchain technology: Beyond bitcoin". In: *Applied Innovation* 2.6-10 (2016), p. 71.

[CTF18] Bihuan Chen, Zhixiong Tan, and Wei Fang. "Blockchain-based implementation for financial product management". In: *2018 28th International Telecommunication Networks and Applications Conference (ITNAC)*. IEEE. 2018, pp. 1–3.

[Cur+17] Vasa Curcin et al. "Templates as a method for implementing data provenance in decision support systems". In: *Journal of biomedical informatics* 65 (2017), pp. 1–21.

[Dag+18] Gaby G Dagher et al. "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology". In: *Sustainable cities and society* 39 (2018), pp. 283–297.

[DL] Wei Dai and Cryptography Mailing List. "Bitcoin Whitepaper". In: ().

[DLS19] Tobias Dehling, Sebastian Lins, and Ali Sunyaev. "Security of critical information infrastructures". In: *Information Technology for Peace and Security*. Springer, 2019, pp. 319–339.

[DMH17] Vikram Dhillon, David Metcalf, and Max Hooper. "The hyperledger project". In: *Blockchain enabled applications*. Springer, 2017, pp. 139–149.

[Eng17] Mark A Engelhardt. "Hitching healthcare to the chain: An introduction to blockchain technology in the healthcare sector". In: *Technology Innovation Management Review* 7.10 (2017).

[Eva+04] Sue M Evans et al. "Anonymity or transparency in reporting of medical error: a community-based survey in South Australia". In: *Medical Journal of Australia* 180.11 (2004), pp. 577–580.

[Eys01] G. Eysenbach. "What is e-health?" In: *J Med Internet Res* 3.2 (June 2001), e20. ISSN: 1438-8871. DOI: 10.2196/jmir.3.2.e20. URL: http://www.ncbi.nlm.nih.gov/pubmed/11720962.

[Fea+12] Nicola T Fear et al. "Does anonymity increase the reporting of mental health symptoms?" In: *BMC public health* 12.1 (2012), pp. 1–7.

[FHS17] Nikolaus Forgó, Stefanie Hänold, and Benjamin Schütze. "The principle of purpose limitation and big data". In: *New technology, big data and the law*. Springer, 2017, pp. 17–42.

[Fog20] Massimo Foglia. "Patients and Privacy: GDPR compliance for healthcare organizations". In: *Eur. J. Privacy L. & Tech.* (2020), p. 43.

[Fos+20] Samuel Fosso Wamba et al. "Bitcoin, blockchain and fintech: a systematic review and case studies in the supply chain". In: *Production Planning & Control* 31.2-3 (2020), pp. 115–142.

[Fre+08] Juliana Freire et al. "Provenance for computational tasks: A survey". In: *Computing in Science & Engineering* 10.3 (2008), pp. 11–21.

[GD07] B. Glavic and K. R. Dittrich. "Data provenance: A Cctegorization of existing approaches". eng. In: *BTW '07: Datenbanksysteme in Buisness, Technologie und Web* 103 (Mar. 2007). Ed. by A. Kemper et al. Conference Name: 12. Fachtagung des GI-Fachbereichs "Datenbanken und Informationssysteme" ISBN: 9783885791973 Meeting Name: 12. Fachtagung des GI-Fachbereichs "Datenbanken und Informationssysteme" Number: 103 Place: Bonn Publisher: Gesellschaft für Informatik (GI), pp. 227–241. DOI: 10.5167/uzh-24450. URL: http://www.btw2007.de/paper/p227.pdf (visited on 2021-06-05).

[Goo02] Susan Dorr Goold. "Trust, distrust and trustworthiness: Lessons from the field". In: *Journal of General Internal Medicine* 17.1 (2002), p. 79.

[Gui12] Luigi Guiso. "Trust and Insurance Markets 1". In: *Economic Notes* 41.1-2 (2012), pp. 1–26.

[Has+20] Anton Hasselgren et al. "Blockchain in healthcare and health sciences—A scoping review". In: *International Journal of Medical Informatics* 134 (2020), p. 104040.

[HB17] Steffen Hoernig and Marc Bourreau. *Interoperability of mobile money International experience and recommendations for Mozambique*. 2017.

[HE18] Ijazul Haq and Olivier Muselemu Esuka. "Blockchain technology in pharmaceutical industry to prevent counterfeit drugs". In: *International Journal of Computer Applications* 180.25 (2018), pp. 8–12.

[Hed08] Hans Hedbom. "A survey on transparency tools for enhancing privacy". In: *IFIP Summer School on the Future of Identity in the Information Society*. Springer. 2008, pp. 67–82.

[HK21] Taylor Hardin and David Kotz. "Amanuensis: Information provenance for health-data systems". In: *Information Processing & Management* 58.2 (2021), p. 102460.

[Höl+18] Marko Hölbl et al. "A systematic review of the use of blockchain in healthcare". In: *Symmetry* 10.10 (2018), p. 470.

[HPH11] Hans Hedbom, Tobias Pulls, and Marit Hansen. "Transparency tools". In: *Privacy and Identity Management for Life*. Springer, 2011, pp. 135–143.

[Hu+20] Rui Hu et al. "A survey on data provenance in IoT". In: *World Wide Web* 23.2 (2020), pp. 1441–1463.

[Hus+21] Muzammil Hussain et al. "Security and Privacy in FinTech: A Policy Enforcement Framework". In: *Research Anthology on Concepts, Applications, and Challenges of FinTech*. IGI Global, 2021, pp. 372–384.

[JRB19] Sandra Johnson, Peter Robinson, and John Brainard. "Sidechains and interoperability". In: *arXiv preprint arXiv:1903.04077* (2019).

[JSZ07] Yong Jin, Wei Song, and Jingyi Zhang. "On Developing China's Third Party Payment". In: *Integration and Innovation Orient to E-Society Volume 1*. Springer, 2007, pp. 578–585.

[Kan+20] Niclas Kannengießer et al. "Trade-offs between distributed ledger technology characteristics". In: *ACM Computing Surveys (CSUR)* 53.2 (2020), pp. 1–37.

[Kee+07] Staffs Keele et al. *Guidelines for performing systematic literature reviews in software engineering*. Tech. rep. Citeseer, 2007.

[KLG03] Spyros Kokolakis, Costas Lambrinoudakis, and Dimitris Gritzalis. "A knowledge-based repository model for security policies management". In: *International Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security*. Springer. 2003, pp. 112–121.

[KMR19] Michelle Krahe, Eleanor Milligan, and Sheena Reilly. "Personal health information in research: perceived risk, trustworthiness and opinions from patients attending a tertiary healthcare facility". In: *Journal of biomedical informatics* 95 (2019), p. 103222.

[KS18] Kevin Klein and Pieter Stolk. "Challenges and opportunities for the traceability of (biological) medicinal products". In: *Drug safety* 41.10 (2018), pp. 911–918.

[KWC18] Maged N Kamel Boulos, James T Wilson, and Kevin A Clauson. *Geospatial blockchain: promises, challenges, and scenarios in health and healthcare*. 2018.

[LAC19] Gary Leeming, John Ainsworth, and David A Clifton. "Blockchain in health care: hype, trust, and digital health". In: *The Lancet* 393.10190 (2019), pp. 2476–2477.

[Le 18] Tran Le Nguyen. "Blockchain in Healthcare: A New Technology Benefit for Both Patients and Doctors". In: *2018 Portland International Conference on Management of Engineering and Technology (PICMET)*. 2018, pp. 1–6. DOI: 10.23919/PICMET.2018.8481969.

[Lee+13] Kisung Lee et al. "Spatio-temporal provenance: Identifying location information from unstructured text". In: *2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*. IEEE. 2013, pp. 499–504.

[Lia+17] Xueping Liang et al. "Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability". In: *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*. IEEE. 2017, pp. 468–477.

[Lil06] Erik Liljegren. "Usability in a medical technology context assessment of methods for usability evaluation of medical equipment". In: *International Journal of Industrial Ergonomics* 36.4 (2006), pp. 345–352.

[Liu+19] Xiyao Liu et al. "A novel robust reversible watermarking scheme for protecting authenticity and integrity of medical images". In: *IEEE Access* 7 (2019), pp. 76580–76598.

[Liu+21] Wei Liu et al. "A donation tracing blockchain model using improved DPoS consensus algorithm". In: *Peer-to-Peer Networking and Applications* (2021), pp. 1–12.

[LJ09] Guoling Lao and Shanshan Jiang. "Risk analysis of third-party online payment based on PEST model". In: *2009 International Conference on Management and Service Science*. IEEE. 2009, pp. 1–5.

[Lov08] Christian Lovis. "Traceability in healthcare: crossing boundaries". In: *Yearbook of medical informatics* 17.01 (2008), pp. 105–113.

[Ma+19] Chaoqun Ma et al. "The privacy protection mechanism of Hyperledger Fabric and its application in supply chain finance". In: *Cybersecurity* 2.1 (2019), pp. 1–9.

[Mam+18] Polina Mamoshina et al. "Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare". In: *Oncotarget* 9.5 (2018), p. 5665.

[Man+18] Suruchi Mann et al. "Blockchain technology for supply chain traceability, transparency and data provenance". In: *Proceedings of the 2018 International Conference on Blockchain Technology and Application*. 2018, pp. 22–26.

[Mar+20] Andrea Margheri et al. "Decentralised provenance for healthcare data". en. In: *International Journal of Medical Informatics* 141 (Sept. 2020), p. 104197. ISSN: 1386-5056. DOI: 10.1016/j.ijmedinf.2020.104197. URL: https://www.sciencedirect.com/science/article/pii/S1386505619312031 (visited on 2021-06-05).

[Mau+17] Roger Maull et al. "Distributed ledger technology: Applications and implications". In: *Strategic Change* 26.5 (2017), pp. 481–489.

[Mos+98] Farzad Mostashari et al. "Acceptance and adherence with antiretroviral therapy among HIV-infected women in a correctional facility." In: *Journal of acquired immune deficiency syndromes and human retrovirology: official publication of the International Retrovirology Association* 18.4 (1998), pp. 341–348.

[MS15] Michael Mainelli and Mike Smith. "Sharing ledgers for sharing economies: an exploration of mutual distributed ledgers (aka blockchain technology)". In: *Journal of financial perspectives* 3.3 (2015).

[Muh14] Jill C Muhrer. "The importance of the history and physical in diagnosis". In: *The Nurse Practitioner* 39.4 (2014), pp. 30–35.

[Mun+06] Kiran-Kumar Muniswamy-Reddy et al. "Provenance-aware storage systems." In: *Usenix annual technical conference, general track*. 2006, pp. 43–56.

[MV08] Mohammad Mannan and Paul C Van Oorschot. "Security and usability: the gap in real-world online banking". In: *Proceedings of the 2007 Workshop on New Security Paradigms*. 2008, pp. 1–14.

[NG17] Christopher Natoli and Vincent Gramoli. "The balance attack or why forkable blockchains are ill-suited for consortium". In: *2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*. IEEE. 2017, pp. 579–590.

[Nor09] Donald A Norman. "THE WAY I SEE IT When security gets in the way". In: *interactions* 16.6 (2009), pp. 60–63.

[Pan+20] Abhishek Kumar Pandey et al. "Key issues in healthcare data integrity: Analysis and recommendations". In: *IEEE Access* 8 (2020), pp. 40612–40628.

[Pet+04] Sandra Petronio et al. "Family and friends as healthcare advocates: Dilemmas of confidentiality and privacy". In: *Journal of Social and Personal Relationships* 21.1 (2004), pp. 33–52.

[Pie+18] Alessandra Pieroni et al. "Smarter city: smart energy grid based on blockchain technology". In: *Int. J. Adv. Sci. Eng. Inf. Technol* 8.1 (2018), pp. 298–306.

[PMV12] M Poulymenopoulou, Flora Malamateniou, and George Vassilacopoulos. "Emergency healthcare process automation using mobile computing and cloud services". In: *Journal of medical systems* 36.5 (2012), pp. 3233–3241.

[Rah+20] Mohamed Abdur Rahman et al. "Secure and provenance enhanced Internet of health things framework: A blockchain managed federated learning approach". In: *Ieee Access* 8 (2020), pp. 205071–205087.

[Rai14] Stephen A Rains. "The implications of stigma and anonymity for self-disclosure in health blogs". In: *Health communication* 29.1 (2014), pp. 23–31.

[Ram+18] Vidhya Ramani et al. "Secure and efficient data accessibility in blockchain based healthcare systems". In: *2018 IEEE Global Communications Conference (GLOBECOM)*. IEEE. 2018, pp. 206–212.

[RDR18] Juan M Roman-Belmonte, Hortensia De la Corte-Rodriguez, and E Carlos Rodriguez-Merchan. "How blockchain technology can change medicine". In: *Postgraduate medicine* 130.4 (2018), pp. 420–427.

[RH13] Fergal Reid and Martin Harrigan. "An analysis of anonymity in the bitcoin system". In: *Security and privacy in social networks*. Springer, 2013, pp. 197–223.

[Rin97a] Thomas C Rindfleisch. "Privacy, information technology, and health care". In: *Communications of the ACM* 40.8 (1997), pp. 92–100.

[Rin97b] Thomas C. Rindfleisch. "Privacy, Information Technology, and Health Care". In: *Commun. ACM* 40.8 (Aug. 1997), pp. 92–100. ISSN: 0001-0782. DOI: 10.1145/257874.257896. URL: https://doi.org/10.1145/257874.257896.

[RK+17] Aravind Ramachandran, Dr Kantarcioglu, et al. "Using blockchain and smart contracts for secure data provenance management". In: *arXiv preprint arXiv:1709.10000* (2017).

[RL18] Igor Radanović and Robert Likić. "Opportunities for use of blockchain technology in medicine". In: *Applied health economics and health policy* 16.5 (2018), pp. 583–590.

[Rob20] Peter Robinson. "The merits of using ethereum mainnet as a coordination blockchain for ethereum private sidechains". In: *The Knowledge Engineering Review* 35 (2020).

[Rog02] Wendy A Rogers. "Is there a moral duty for doctors to trust patients?" In: *Journal of Medical Ethics* 28.2 (2002), pp. 77–80.

[SAD19] Hadi Saleh, Sergey Avdoshin, and Azamat Dzhonov. "Platform for tracking donations of charitable foundations based on blockchain technology". In: *2019 Actual Problems of Systems and Software Engineering (APSSE)*. IEEE. 2019, pp. 182–187.

[Saf+98] Dana Gelb Safran et al. "Linking primary care performance to outcomes of care". In: *Journal of family practice* 47 (1998), pp. 213–220.

[Sca16] Claudio Scardovi. *Restructuring and innovation in banking*. Springer, 2016.

[SDS19] Markus Schäffer, Monika Di Angelo, and Gernot Salzer. "Performance and scalability of private Ethereum blockchains". In: *International Conference on Business Process Management*. Springer. 2019, pp. 103–118.

[Sen] Oshani W Seneviratne. "Data Provenance and Accountability on the Web". In: *Provenance in Data Science: From Data Models to Context-Aware Knowledge Graphs* (), p. 11.

[Sin+20] Aashutosh Singh et al. "Aid, Charity and donation tracking system using blockchain". In: *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*. IEEE. 2020, pp. 457–462.

[Sir+19] N Sai Sirisha et al. "Proposed solution for trackable donations using blockchain". In: *2019 International Conference on Nascent Technologies in Engineering (ICNTE)*. IEEE. 2019, pp. 1–5.

[SMK09] John B Smelcer, Hal Miller-Jacobs, and Lyle Kantrovich. "Usability of electronic medical records". In: *Journal of usability studies* 4.2 (2009), pp. 70–84.

[SPG05] Yogesh L Simmhan, Beth Plale, and Dennis Gannon. "A survey of data provenance in e-science". In: *ACM Sigmod Record* 34.3 (2005), pp. 31–36.

[SR09] Mats Skoglund and Per Runeson. "Reference-based search strategies in systematic reviews". In: *13th International Conference on Evaluation and Assessment in Software Engineering (EASE) 13*. 2009, pp. 1–10.

[SS17] Andreas Schreiber and Regina Struminski. "Tracing personal data using comics". In: *International Conference on Universal Access in Human-Computer Interaction*. Springer. 2017, pp. 444–455.

[SSS18] P Sajana, M Sindhu, and M Sethumadhavan. "On blockchain applications: hyperledger fabric and ethereum". In: *International Journal of Pure and Applied Mathematics* 118.18 (2018), pp. 2965–2970.

[Sun+14] Ali Sunyaev et al. "Availability and quality of mobile health app privacy policies". In: *Journal of the American Medical Informatics Association* 22.e1 (Aug. 2014), e28–e33. ISSN: 1067-5027. DOI: 10.1136/amiajnl-2013-002605. eprint: https://academic.oup.com/jamia/article-pdf/22/e1/e28/34145987/amiajnl-2013-002605.pdf. URL: https://doi.org/10.1136/amiajnl-2013-002605.

[Sun20] Ali Sunyaev. "Distributed ledger technology". In: *Internet Computing*. Springer, 2020, pp. 265–299.

[SVK12] Frédérique Six, Marianne van der Veen, and Niels Kruithof. "Conceptualizing integrity systems in governments and banking". In: *Public Integrity* 14.4 (2012), pp. 361–382.

[Swa15] Melanie Swan. *Blockchain: Blueprint for a new economy*. " O'Reilly Media, Inc.", 2015.

[SWR97] Dennis D Steinauer, Shukri A Wakid, and Stanley Rasberry. "Trust and traceability in electronic commerce". In: *StandardView* 5.3 (1997), pp. 118–124.

[TBA16] Yucel Tas, Mohamed Jehad Baeth, and Mehmet S Aktas. "An approach to standalone provenance systems for big social provenance data". In: *2016 12th International Conference on Semantics, Knowledge and Grids (SKG)*. IEEE. 2016, pp. 9–16.

[Tei18] Fabian Maximilian Johannes Teichmann. "Financing terrorism through cryptocurrencies–a danger for Europe?" In: *Journal of Money Laundering Control* (2018).

[TQV21] Ofir Turel, Hamed Qahri-Saremi, and Isaac Vaghefi. "Special Issue: Dark Sides of Digitalization". In: *International Journal of Electronic Commerce* 25.2 (2021), pp. 127–135. DOI: 10.1080/10864415.2021.1887694. eprint: https://doi.org/10.1080/10864415.2021.1887694. URL: https://doi.org/10.1080/10864415.2021.1887694.

[Tru18] TrueVault. *What is personal data? - TrueVault*. en. Nov. 2018. URL: https://www.truevault.com/learn/what-is-personal-data (visited on 2021-06-05).

[Tsa+07] Wei-Tek Tsai et al. "Data provenance in SOA: security, reliability, and integrity". In: *Service Oriented Computing and Applications* 1.4 (2007), pp. 223–247.

[Udd+18] Md Ashraf Uddin et al. "Continuous patient monitoring with a patient centric agent: A block architecture". In: *IEEE Access* 6 (2018), pp. 32700–32726.

[Ung+06] Brigitte Unger et al. "The amounts and the effects of money laundering". In: *Report for the Ministry of Finance* 16.2020.08 (2006), p. 22.

[Vuk15] Marko Vukolić. "The quest for scalable blockchain fabric: Proof-of-work vs. BFT replication". In: *International workshop on open problems in network security*. Springer. 2015, pp. 112–125.

[Wal06] Tom Walley. *Using personal health information in medical research*. 2006.

[War+04] Hester JT Ward et al. "Obstacles to conducting epidemiological research in the UK general population". In: *bmj* 329.7460 (2004), pp. 277–279.

[WN94a] Claes-Goran Westrin and Tore Nilstun. "The ethics of data utilisation: a comparison between epidemiology and journalism". In: *BMJ* 308.6927 (1994), pp. 522–523.

[WN94b] Claes-Goran Westrin and Tore Nilstun. "The ethics of data utilisation: a comparison between epidemiology and journalism". In: *BMJ* 308.6927 (1994), pp. 522–523.

[Wor+20] Carl Worley et al. "Scrybe: A Second-Generation Blockchain Technology with Lightweight Mining for Secure Provenance and Related". In: *Blockchain Cybersecurity, Trust and Privacy* 79 (2020), p. 51.

[WW02] Jane Webster and Richard T Watson. "Analyzing the past to prepare for the future: Writing a literature review". In: *MIS quarterly* (2002), pp. xiii–xxiii.

[XCK19] Min Xu, Xingtong Chen, and Gang Kou. "A systematic review of blockchain". In: *Financial Innovation* 5.1 (2019), pp. 1–14.

[Xia+17] QI Xia et al. "MeDShare: Trust-less medical data sharing among cloud service providers via blockchain". In: *IEEE Access* 5 (2017), pp. 14757–14767.

[Xu+17] Xiwei Xu et al. "A taxonomy of blockchain-based systems for architecture design". In: *2017 IEEE international conference on software architecture (ICSA)*. IEEE. 2017, pp. 243–252.

[Yan17] Hyoeun Yang. "The UK's Fintech Industry Support Policies and its Implications". In: *KIEP Research Paper, World Economy Brief* (2017), pp. 17–05.

[Yeo+17] Kimchai Yeow et al. "Decentralized consensus for edge-centric internet of things: A review, taxonomy, and research issues". In: *IEEE Access* 6 (2017), pp. 1513–1524.

[Yli+16] Jesse Yli-Huumo et al. "Where is current research on blockchain technology?—a systematic review". In: *PloS one* 11.10 (2016), e0163477.

[Yoo17] Soonduck Yoo. "Blockchain based financial case analysis and its implications". In: *Asia Pacific Journal of Innovation and Entrepreneurship* (2017).

[Zha+18a] Peng Zhang et al. "FHIRChain: applying blockchain to securely and scalably share clinical data". In: *Computational and structural biotechnology journal* 16 (2018), pp. 267–278.

[Zha+18b] Huawei Zhao et al. "Efficient key management scheme for health blockchain". In: *CAAI Transactions on Intelligence Technology* 3.2 (2018), pp. 114–118.

[ZJ18] Kaiwen Zhang and Hans-Arno Jacobsen. "Towards Dependable, Scalable, and Pervasive Distributed Ledgers with Blockchains (Technical Report)". In: (2018).

[ZN+15] Guy Zyskind, Oz Nathan, et al. "Decentralizing privacy: Using blockchain to protect personal data". In: *2015 IEEE Security and Privacy Workshops*. IEEE. 2015, pp. 180–184.

[Zub15] Shoshana Zuboff. "Big other: surveillance capitalism and the prospects of an information civilization". In: *Journal of Information Technology* 30.1 (Mar. 2015), pp. 75–89. ISSN: 1466-4437. DOI: 10.1057/jit.2015.5. URL: https://doi.org/10.1057/jit.2015.5.