Karlsruhe Institute of Technology

DSN

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

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 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 that are mutually intertwined. 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?

The scope of this stage is limited to literature on (1) DLTs in the domain of (2) healthcare and finance.

Q4: Which of the DP requirements are fulfilled/unfulfilled or 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.

Our selections of search strings were based on some commonly used terms and acronyms, such as "provenance", "lineage", "traceability", "transparency", etc. ; "health", "healthcare", "medical", "biomedical", "clinical", etc. as well as "EHR" (Electronic Health Records), "EMR" (Electronic Medical Records) and "PHR" (Personal Health Records); "finance", "money", "banking", "fintech", etc. Although different by definition, switching between "blockchain" and "distributed ledger technology" didn't provide any different results in our case.

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.

1.4 Outline

In the next section, take a closer look at data provenance, the requirements of such approaches and the use cases selected in our work. In section three we describe distributed ledger technologies, their different designs, characteristics and properties, as well as DLTs' suitability for DP. Section four presents an evaluated mapping of our selected DLT approaches to the financial and healthcare DP requirements. This is followed by discussion in section five, consisting of principle findings, implications for practice, implications for research, limitations and future work. Then we end the work with a brief conclusion in section six.

2 Data Provenance

2.1 Definition

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. 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 Requirements

DP approaches/technologies, suitable for tracing the origin and source of personal data and the processes that led to its current state, have numerous requirements. By analysing the relevant literature we found on DP [table with all DP papers we derived req from], 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 accountability, as well as pseudonymity, anonymity and unlinkability. Security is defined by confidentiality, integrity and availability. 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.).

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	By definition - the capability to communicate, execute programs or transfer data between various systems in a manner that requires Data Subjects to have little or no knowledge of the unique characteristics of those systems [, 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].

2.3 Use Cases

In this work we investigate personal DP approaches for both *healthcare* and *finance*. While all of the above mentioned requirements are important in such techlonogies, each requirement can have different level of significance and meaning in the specific use case. In this section, we will discuss these differences, specifics and nuances.

2.3.1 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].

2.3.2 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. Finally, it is 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].

3 Distributed Ledger Technologies

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.

3.1 Designs, Properties and Characteristics

3.1.1 Designs

DLT designs can be instantiated as a *public* or *private* [Xu+17a; 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].

Besides the choice of going with *public* or *private*, we differentiate between *permissioned* and *permissionless* DLT designs [Yeo+17; Xu+17a].

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].

3.1.2 Properties

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 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

3.1.3 Characteristics

Out of those 40 DLT characteristics, we identified 10 "main" and 13 "secondary" characteristics, which correspond to our DP approach requirements described in table 2.2. These 23 characteristics can help us determine whether DLT is suitable for DP and a potential solution some of its present issues, or not.

Main 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
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
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
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
Turing-complete Smart Contracts	The support of Turing-complete smart contracts within a DLT design
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
Strength of Cryptography	The difficulty of breaking the cryptographic algorithms used in the DLT design
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
Ease of Node Setup	The ease of configuring and adding a new (or previously crashed) node to the distributed ledger

3.2 DLT and DP

One of DLTs' properties 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].

3.3 DLT in Healthcare

3.4 DLT in Finance

3.5 Considered Approaches

3.5.1 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].

3.5.2 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.

3.5.3 Comparison

From Hyperledger Fabric's and Ethereum's white papers it becomes clear that both systems have differing views on future implementation areas. By comparison, Fabric aims to provide a comprehensive and extensible module architecture that can be used in various industries

(finance, healthcare, supply chain). However, Ethereum is totally outside of every particular area of application and serves more as a general interface for all types of transactions and applications. Ethereum executes random complexity codes through its EVM and is freely available without authorization to any person. Fabric has a scalable design that guarantees durability and hence scalability with an allowed operating mode. Smart contract codes are used by both implementations. The chaincode in Fabric is known and unique consumer networks exist where the communications and relevant transactions of the associated channels can be seen. Access to transactions that offer anonymity to transactions is also limited. Both use various consensus processes to make the conclusion. Ethereum has Ether, but Fabric would not need an integrated currency since there is no mining. Ethereum has a generic platform that runs powerful smart contracts which is public and transparent and Fabric's modular architecture allows a customized platform for a specific mode of operation [MAA21].

Apart from healthcare and finance, which we will discuss later, Ethereum has been used in a variety of domains, including: energy [], IoT [], naming service [], art [], gaming [] and crowd-funding (DAO and Initial Coin Offerings).

However, public DLT designs lack speed, face concerns over scalability, energy consumption, privacy, identity and cost [Cho+19]. The bigger the public network, the slower it is, as more transactions take place and clog the network. For example, Ethereum can handle 13 transactions per second, compared to 3000 by Hyperledger Fabric, compared to 56 000 by VISA. Furthermore, data stored in these ledgers are visible to any participant and therefore not suitable to handle sensitive data. All entities are identified via cryptographic pseudonyms that makes it hard to audit and are open to Sybil attacks [Dou02]. Storing data also incurs expense and is therefore infeasible to store a large amount of data in the ledger. 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. Public blockchain is also often criticized for lack of governance when it comes to cyber crime [Why19], money laundering [MBB13] and drug trading [Heg16].

Private DLT approaches have been created to resolve the issues of public DLT designs. They do not rely on an energy- intensive consensus algorithm. They can process transactions much faster than any public DLT system. The identity of each entity within the network is verified, thereby enabling auditability and accountability. This also acts as a mechanism for protection against Sybil attacks. Private DLT systems afford federated access to data in the ledger, ensuring privacy by only allowing authorized entities to access any private and sensitive data in the system. Fabric, for example, supports the upgradability of smart contracts, which can be vital when new logic needs to be incorporated or a bug needs to be fixed. All these properties make the private DLT systems suitable for scenarios that need to deal with highly sensitive data.

However, the biggest disadvantages of private DLT designs is centralization. Private distributed ledgers inherently become centralized due to their private network. 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. Private DLT systems cannot provide the same amount of security regarding the immutability of data and code in comparison to public DLT systems.

The reason we consider these DLT approaches is that many of the Ethereum's (public and private) and Hyperledger Fabric's complementing features overlap with our defined DP requirements (table 2.2). We will further discuss these relations in section 4.2.

...

3.5.4 Trade-Offs

From N. Kannengießer et al. [Kan+20]:

A	B	***Ethereum***	***H. Fabric***
Confidentiality (+)	Integrity (+)	B	A
Consistency (+)	Availability (+)	B	A
Strength of Cryptography (+)	Support for Constr. Devices (+)	A	-
Maintainability (+)	Availability (+)	B	A
	Integrity (+)	B	A
Node Controller Verification (+)	Ease of Node Setup (+)	B	A
Turing-compl. Smart Contracts (+)	Confidentiality (+)	A	AB
Throughput (+)	Consistency (+)	B	A
	Integrity (+)	B	A
User Unidentifiability (+)	Throughput (+)	A	B

A: *DLT characteristic A out-weights DLT characteristic B*
B: *DLT characteristic B out-weights DLT characteristic A*
AB: *DLT characteristic A and B are both achieved (trade-off avoided by other means)*
-: *Trade-off not applicable*
(+): *DLT characteristic is aimed to be high*

Confidentiality vs **Integrity**
To improve confidentiality, DLT designs are often implemented in a private network, where only select nodes can join (private DLT designs like, for example, a private Ethereum or Hyperledger Fabric blockchain). However, a small number of known nodes makes it easier to have detailed information on the network topology. Access to a detailed network topology facilitates initiation of targeted delays in the communication between nodes, because the data flow is known [NG17]. Thus, 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].

Consistency vs **Availability**
Distributed systems theory reveals a trade-off between consistency and availability—the CAP Theorem [Aba12]. This trade-off also persists in the field of DLT and is caused by latency in block propagation, for example, due to big block sizes or network failures. The larger the number of nodes that must receive new transactions, the longer the distributed ledger is in an inconsistent state. The larger the number of nodes of a distributed ledger, the more time it takes until each node has received the new block. However, many replications of the data stored on the distributed ledger increases availability. Thus, there is a trade-off between high availability and fast consistency [Kan+20].

Strength of Cryptography vs **Support for Constrained Devices**
Low collision [] likelihood is desirable, which is why more secure hashing and key generation approaches are required (e.g., more bits for the hash). However, an increased strength of cryptography requires more computational resources, such as random access memory and storage memory [Mal+16]. Thus, constrained devices such as microcontrollers can hardly handle resource- intensive cryptography [Mal+16; KBL18].

Maintainability vs **Availability**
To secure DLT designs, the software client of individual nodes must be maintainable and remain compatible with the majority of nodes in the network. Updates of the client protocol of a DLT design must be performed on each node. This is why maintainability of DLT designs decreases with an increasing number of independent nodes due to additional efforts when negotiating and applying software client updates. However, an increasing number of nodes increases the ledger's redundancy due to increasing replications. The dependency between maintenance-related cost (e.g. time and money) and the degree of decentralization of the distributed ledger is also derogatorily refered to as blockchain bloat or DLT bloat [Put+18; Kan+20].

Maintainability vs **Integrity**
To allow for efficient maintenance of a distributed ledger, the coordination of update procedures should be facilitated by a low number of (independently controlled) nodes. However, a decrease in the number of independently maintained nodes (hence, a decrease in the DLT design's degree of decentralization) impedes the integrity of DLT designs due to reduced absolute fault tolerance regarding the number of tolerable, malicious nodes.

However, a high level of a distributed ledger's integrity also impacts maintainability of applications on DLT [Cob17]. For achieving a high integrity of distributed ledgers, smart contracts are tamper-resistant: Smart contracts must always be redeployed and initialized with the state of the obsolete version whenever the smart contract should be updated. In addition, the new address of the updated smart contract must be adapted in any module of the corresponding applications and the chained smart contracts that reference the deprecated smart contract. Hence, tamper-resistance and resulting integrity of a distributed ledger increases efforts for maintenance. However, by relaxing integrity, thus, tamper-resistance of smart contracts, the idea of an inevitable and automated enforcement of agreements becomes vulnerable to malicious behavior [Kan+20].

Node Controller Verification vs **Ease of Node Setup**

Verification of node controllers and their node permissions (e.g. permissions to read data or to validate and commit new transactions) is required in permissioned DLT designs [68]. After permissions are granted to a node, the node can participate in (mostly, voting-based) consensus mechanisms. A public key infrastructure (PKI) with a trusted certification authority is often used to verify the nodes' identities and issue certificates to nodes [Hof18]. However, a PKI produces additional efforts to obtain a certificate for the public-private key pair and leads to the dependency on a trusted certificate authority. Consequently, it becomes more complex to set up a node and participate in a permissioned distributed ledger compared to public-permissionless DLT designs (like Ethereum), which only require to install an open-source software client (e.g. Geth or Parity for Ethereum) [Kan+20].

Turing-complete Smart Contracts vs **Confidentiality**

Use of smart contracts threatens confidentiality in three ways. First, it is publicly visible which account's transactions triggered a smart contract [Kos+16]. Second, the compiled smart contract code is also visible to the public and smart contracts can be decompiled to human readable source code. Thus, the current state of the smart contract and even values of variables that are declared private in the smart contract can be inferred due to the open smart contract code and transactions [Kos+16]. Hence, the common ways of using smart contracts do not support confidentiality. Nevertheless, there are new approaches for private smart contracts. For example, smart contracts can be divided into a private and a public part. The private part determines the payout distribution among involved parties; the input data (e.g. a number of coins) is kept private and is protected using zero-knowledge proofs [Kos+16]. As a result, no participant knows the input data other participants sent to the smart contract. Third, oracles and external services might have insight into data that are exchanged via smart contracts. Such oracles or services are often centralized instances that forward certain data. The use of external services in DLT requires at least one trusted party that stores the requested data. Thus, the oracle provider and also the provider of the external data feed can have insights into data flows that are made by users who trigger a smart contract [Kan+20].

Throughput vs **Consistency**

For the DLT concept blockchain, it was found that an increased block size can increase throughput, because more transactions can be included in a block [GK17]. An increased block size comes with a longer propagation delay [GK17; GL02; Xu+17b], which results in a longer state of inconsistency between nodes in a distributed ledger [DW13; Unt+18]. For Ethereum, it was found that the percentage of created blocks that are successfully committed to the blockchain's main chain becomes low as the block size (and consequently the block propagation delay) increases [GK17]. Consequently, the stale block rate increases and nodes have inconsistent views on the ledger until the forks are resolved. Such inconsistent states facilitate successful attacks. Forkable DLT designs based on PoW can only improve throughput by degrading consistency, thereby, and increasing vulnerability [Bar+17; Kan+20].

Throughput vs **Integrity**

Increased block size can increase throughput, because more transactions can be included in a block [GK17]. The increased throughput comes with longer block propagation delays, because more transactions are included in a block. However, longer block propagation delays increase the probability of forks [Xu+17b], which threaten integrity and facilitate successful partition-based attacks on the distributed ledger [Hof+17] (e.g. selfish-mining [Ger+16], long-range attacks [DPP19], bribery attacks [Bon16]). To preserve integrity, the block creation interval must be adjusted. Nevertheless, long block creation intervals also decrease the number of blocks issued, ultimately decreasing throughput. An increase of the block size to include more transactions per block will increase the message propagation delay and, thus, the number of forks in the distributed ledger [Kan+20].

User Unidentifiability vs **Throughput**

The less a network is controlled by a central authority and the more nodes participate in the network (given a high degree of decentralization), the more vague is the identity of nodes. Therefore, public-permissionless DLTs promise higher user unidentifiability than permissioned ones due to typically a higher number of nodes and a higher degree of decentralization. In contrast, a smaller, permissioned distributed ledger with verified and identifiable nodes allows for higher throughput, because faster consensus mechanisms can be used. Nevertheless, unidentifiability can be improved by applying additional processes like mixing and the use of new keypairs for each transaction [Zie+15]. Yet, these processes create overhead due to preprocessing of each transaction, which results in extended transaction validation speed and hence decreases throughput [Kan+20].

3.5.5 Suitability for Healthcare DP

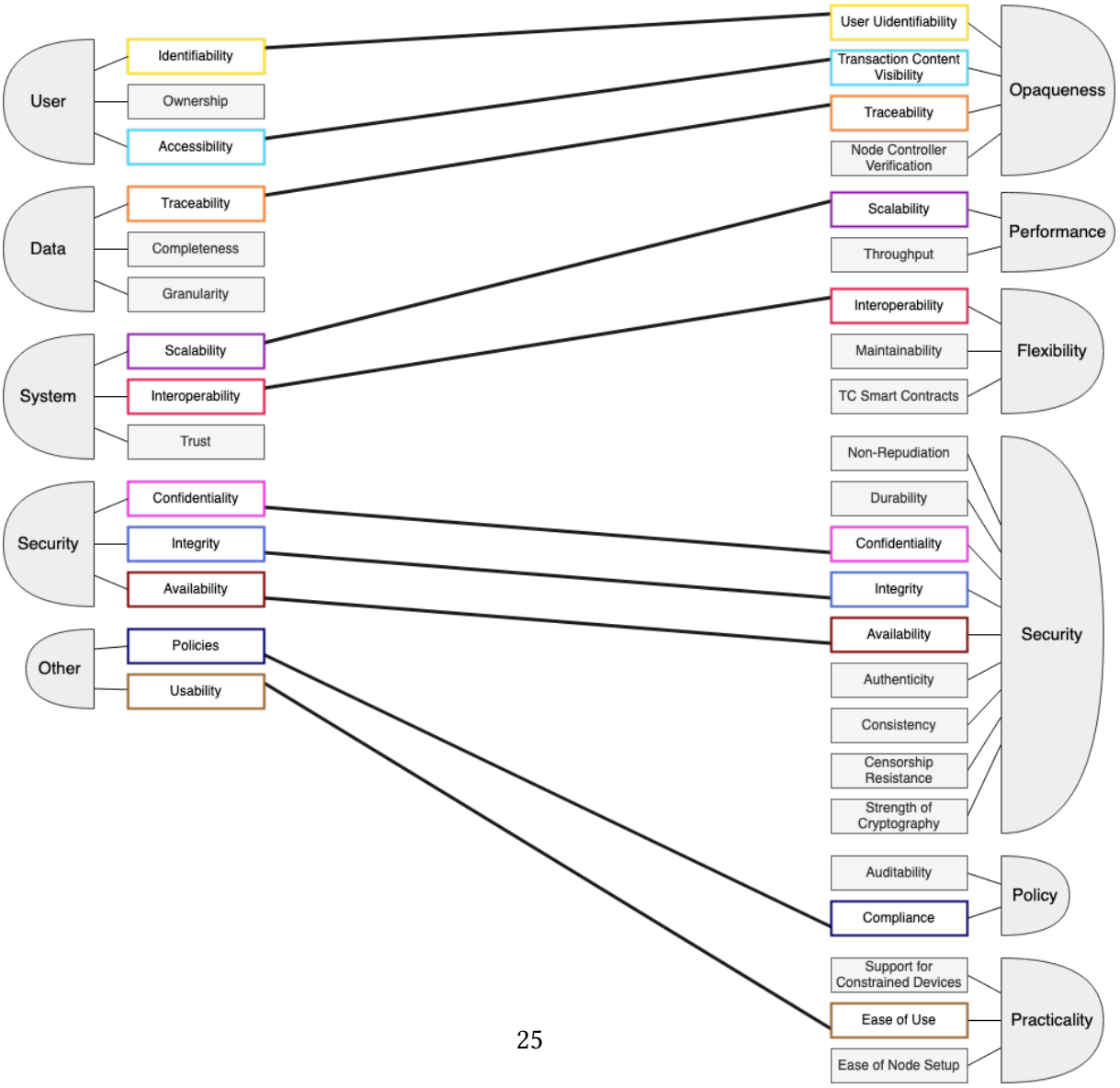
3.5.6 Suitability for Finance DP

4 Evaluated Mapping

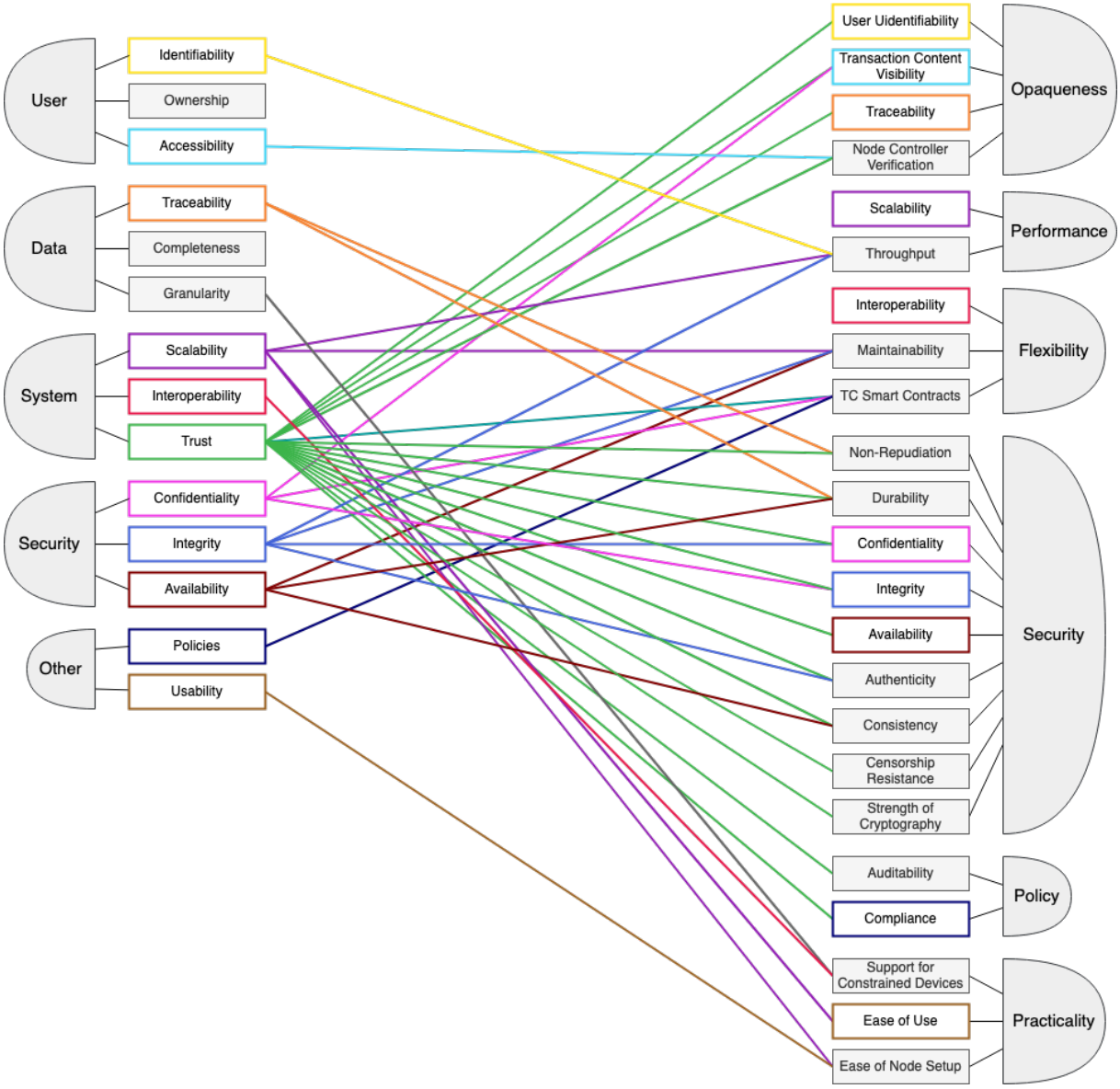
4.1 DP Requirements to DLT Characteristics

The 10 "main" DLT characteristics correspond directly to some of our defined DP requirements. The other 13 "secondary" DLT characteristics that we identified are rather concrete and specific. Furthermore, our defined requirements in table 2.2. can be influenced by one or more DLT characteristics. Here we aim to present in an easily visible way, which DLT characteristics seem to be related to which DP requirement.

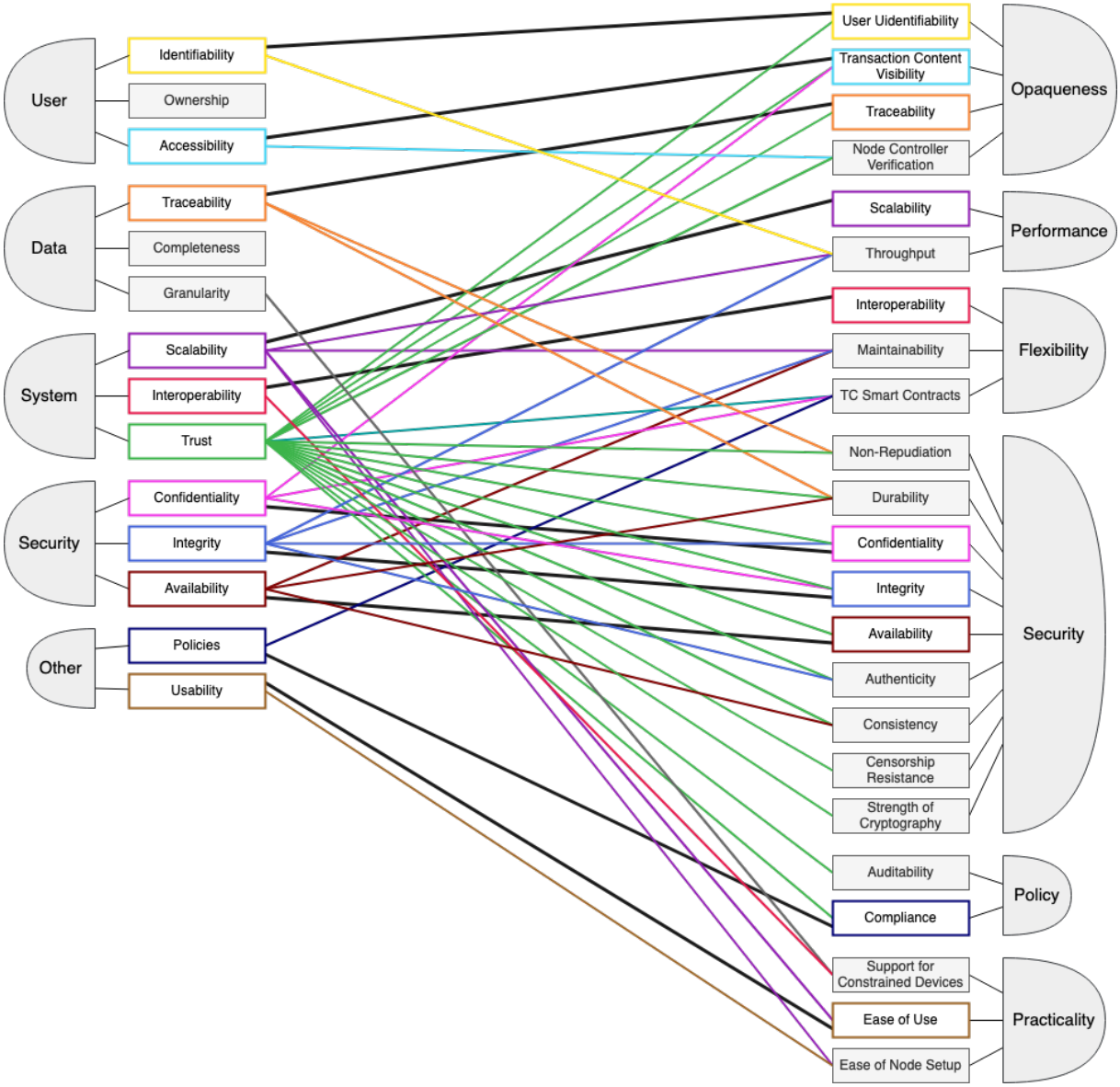
4.1.1 Main Relationships



4.1.2 Secondary Relationships



4.1.3 All Relationships



4.2 DP Requirements to the Considered Approaches' Features

4.2.1 Hyperledger Fabric and DP

Since the network is permissioned, every user participating in a transaction must register in the network for getting their corresponding IDs, which can fulfil our **Identifiability** requirement [SSS18]. Another one of the many compelling Fabric features is the enablement of a network of networks. Members of a network work together, but because businesses need some of their data to remain private, they often maintain separate relationships within their networks [DMH17]. This can mean different **Accesibility** rights for different users. Access rights and the nature of blockchain can provide **Traceability**. Its highly modular and permissioned architecture can bring **Confidentiality**, transparency and accountability which helps achieve high level of **Trust**. Its high throughput (3000 transactions per second) and low latency relate to **Scalability** and can ensure **Availability** in emergency situations. 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. Queryable Data is also a feature which can positively impact **Usability**. **Policies** (and potentially other requirements) can also be implemented through chaincode [DMH17].

4.2.2 Ethereum and DP

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 identification via public key (**Identifiability**). Its immutable computer logic together with smart contracts, EVM, consensus mechanism, Ether (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**. 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**). 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. 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].

In spite of its attractive potential, however, the mass adoption of this technology is prevented by its inherent limitation in scalability and due to Ethereum's probabilistic finality [Vuk15], it is not well suited to information that needs to be available and acted upon immediately [Rob20]. From the scalability perspective, Blockchain-based systems like Ethereum, are commonly compared [Cro+16] 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].

However, 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 trust.

4.2.3 Overview

Requirement	***Hyperledger Fabric***	***Ethereum (public)***	***Ethereum (private)***
Identifiability	x	x	x
Ownership			
Accessibility	x		
Traceability	x	x	x
Completeness			
Granularity			
Scalability	x		x
Interoperability	x		x
Trust	x	x	
Confidentiality	x		x
Integrity		x	
Availability	x	x	x
Policies	x	x	x
Usability	x	x	x

5 Discussion

…

5.1 Principle Findings

…

5.2 Implications for Practice

…

5.3 Implications for Research

…

5.4 Limitations and Future Work

6 Conclusion

…

Bibliography

[] URL: http://www.hashcash.org/papers/announce.txt (visited on 2021-12-05).

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

[] *Blockchains | Unveiling The Future Of Blockchain Technology*. EN. URL: https://www.blockchains.com/ (visited on 2021-12-05).

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

[] *Electron | Empowering Distributed Energy Markets*. en-US. URL: https://electron.net/ (visited on 2021-12-05).

[] *Ethereum Name Service*. URL: https://ens.domains/ (visited on 2021-12-05).

[] *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).

[] *NonFungible.com | Analyze, track & discover NFTs and collectible art markets*. en. URL: https://nonfungible.com (visited on 2021-12-05).

[] *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).

[] *Welcome to Decentraland*. en. URL: https://decentraland.org/ (visited on 2021-12-05).

[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).

[Aba12] Daniel Abadi. "Consistency tradeoffs in modern distributed database system design: CAP is only part of the story". In: *Computer* 45.2 (2012), pp. 37–42.

[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.

[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.

[Bar+17] Artem Barger et al. "Scalable communication middleware for permissioned distributed ledgers". In: *Proceedings of the 10th ACM International Systems and Storage Conference*. 2017, pp. 1–1.

[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.

[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.

[Bon16] Joseph Bonneau. “Why buy when you can rent?” In: *International Conference on Financial Cryptography and Data Security*. Springer. 2016, pp. 19–26.

[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).

[Cob17] Michael Coblenz. "Obsidian: a safer blockchain programming language". In: *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. IEEE. 2017, pp. 97–99.

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

[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.

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

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

[Dou02] John R Douceur. "The sybil attack". In: *International workshop on peer-to-peer systems*. Springer. 2002, pp. 251–260.

[DPP19] Evangelos Deirmentzoglou, Georgios Papakyriakopoulos, and Constantinos Patsakis. "A survey on long-range attacks for proof of stake protocols". In: *IEEE Access* 7 (2019), pp. 28712–28725.

[DW13] Christian Decker and Roger Wattenhofer. "Information propagation in the bitcoin network". In: *IEEE P2P 2013 Proceedings*. IEEE. 2013, pp. 1–10.

[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.

[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).

[Ger+16] Arthur Gervais et al. "On the security and performance of proof of work blockchains". In: *Proceedings of the 2016 ACM SIGSAC conference on computer and communications security*. 2016, pp. 3–16.

[GK17] Johannes Göbel and Anthony E Krzesinski. "Increased block size and Bitcoin blockchain dynamics". In: *2017 27th International Telecommunication Networks and Applications Conference (ITNAC)*. IEEE. 2017, pp. 1–6.

[GL02] Seth Gilbert and Nancy Lynch. "Brewer's conjecture and the feasibility of consistent, available, partition-tolerant web services". In: *Acm Sigact News* 33.2 (2002), pp. 51–59.

[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.

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

[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.

[Heg16] Kartik Hegadekatti. "Regulating the deep Web through controlled blockchains and crypto-currency networks". In: *Available at SSRN 2888744* (2016).

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

[Hof+17] Frank Hofmann et al. "The immutability concept of blockchains and benefits of early standardization". In: *2017 ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K)*. IEEE. 2017, pp. 1–8.

[Hof18] Michał R Hoffman. "Can blockchains and linked data advance taxation". In: *Companion Proceedings of the The Web Conference 2018*. 2018, pp. 1179–1182.

[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.

[KBL18] Djamel Eddine Kouicem, Abdelmadjid Bouabdallah, and Hicham Lakhlef. "Internet of things security: A top-down survey". In: *Computer Networks* 141 (2018), pp. 199–221.

[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.

[Kos+16] Ahmed Kosba et al. "Hawk: The blockchain model of cryptography and privacy-preserving smart contracts". In: *2016 IEEE symposium on security and privacy (SP)*. IEEE. 2016, pp. 839–858.

[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.

[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.

[MAA21] Alaa Hamid Mohammed, Alaa Amjed Abdulateef, and Ihsan Amjad Abdulateef. "Hyperledger, Ethereum and Blockchain Technology: A Short Overview". In: *2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE. 2021, pp. 1–6.

[Mal+16] Lukas Malina et al. "On perspective of security and privacy-preserving solutions in the internet of things". In: *Computer Networks* 102 (2016), pp. 83–95.

[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.

[MBB13] Malte Möser, Rainer Böhme, and Dominic Breuker. "An inquiry into money laundering tools in the Bitcoin ecosystem". In: *2013 APWG eCrime researchers summit*. Ieee. 2013, pp. 1–14.

[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.

[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.

[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.

[Put+18] Deepak Puthal et al. "The blockchain as a decentralized security framework [future directions]". In: *IEEE Consumer Electronics Magazine* 7.2 (2018), pp. 18–21.

[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.

[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.

[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.

[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.

[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.

[Unt+18] Andreas Unterweger et al. "Lessons learned from implementing a privacy-preserving smart contract in ethereum". In: *2018 9th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*. IEEE. 2018, pp. 1–5.

[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.

[Why19] Christopher Whyte. "Cryptoterrorism: Assessing the utility of blockchain technologies for terrorist enterprise". In: *Studies in Conflict & Terrorism* (2019), pp. 1–24.

[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.

[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+17a] 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.

[Xu+17b] 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.

[Zie+15] Jan Henrik Ziegeldorf et al. “Coinparty: Secure multi-party mixing of bitcoins”. In: *Proceedings of the 5th ACM Conference on Data and Application Security and Privacy*. 2015, pp. 75–86.

[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.