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

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, transparent 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].

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. By leveraging the global-scale computing power of distributed networks, a DLT-based DP can provide integrity, authenticity, transparency, accountability, provenance and trustworthiness 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].

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 select the most suitable approach for personal DP in the field of *healthcare* and *finance*. This leads us to the research question: *What are the properties of Distributed Ledger Technologies that make them beneficial/suitable for personal data provenance in healthcare and finance?*

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

In this work we define *data provenance* (DP) as an approach/technology that can be used to record 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 transparent way, that ensures accountability and is in accordance to international laws and regulation, with the well-being of the consumer in mind.

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 to fulfil a number of requirements. Using the available literature, we derived and formulated the following requirements, which we then presented through the lens of the two use cases investigated in our work.

We combined the requirements into suitable groups: "User" contains requirements associated with the data of the individual user; "Data" contains the features that the data is required to posses in such approaches/technologies; "System" consists of requirements about the concrete DP approach/technology as an usable system; "Security" and "Other" include the fundamental security requirements and requirements that are necessary in almost every system.

Group	Requirement	Description
User	Identification	An unique identifier allows identification and lays the ground for accountability [Lee+13], but also anonymity, pseudonymity and unlinkability should be possible [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].
Data	Traceability	Give 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].
	Completeness	Collecting complete provenance information can fully take the advance to track data and actions for identity management, error detection, etc. Incomplete provenance data may lead to detection missing and suppression of abnormal behaviors [GGM12; HPH11].
	Granularity	Not only the process derivation of a data file, but also the components of files such as paragraphs, shapes and images should be traced with regard to their origins. Fine-grained provenance information helps achieve highly precise anomaly detection and auditing [HWA10].
⋮	⋮	⋮

⋮	⋮	⋮
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].
	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].
Security	Confidentiality	Ensures non-disclosure of data traveling over the network to unauthorised Data Subjects [Asg+12].
	Integrity	Ensures that the Data Receiver may detect unauthorised changes made to the data [Tsa+07].
	Availability	Ensuring that data and is provenance is available to Data Subjects,when and where they need it [Lia+17].
Other	Policies	Enforce laws [] 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].
	Logging	Provides mechanisms to log and timestamp the transfer of the data between Data Subjects [HPH11; MZX16; Wan+16; Sue+13].

2.3 Use Cases

In this work we investigate DP approaches for both *healthcare* and *finance*. While such approaches need to fulfil all of the above mentioned requirements, each requirement can have different level of importance or meaning in the specific use case. In this section, we will discuss these differences between *healthcare* and *finance* requirements.

The marked fields in the following table show which requirements are more nuanced than the general DP requirements and, consequently, where the requirements differ in meaning and importance (empty means that the requirement is similar to the general description in table 2.2).

Group	Requirement	*Healthcare*	*Finance*
User	Identification	x	x
	Ownership	x	
	Accessibility	x	x
Data	Traceability	x	x
	Completeness		
	Granularity		
System	Scalability		x
	Interoperability		
	Trust	x	x
Security	Confidentiality	x	x
	Integrity		
	Availability	x	
Other	Policies	x	
	Usability		
	Logging		

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 specific healthcare DP requirements pointed out in table 2.3.

Identification: 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].

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].

Confidentiality: Trust and confidentiality 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].

Availability: Medical data and its provenance should be available and ready for immediate use, especially in cases of emergency [KLG03]. 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].

Policies: 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].

Logging; system **Usability, Scalability** and **Interoperability**; data **Integrity, Completeness** and **Granularity** are, of course, of great importance and are requirements that should be fulfilled in every DP approach. However, these requirements don't seem to have any specific aspects that differ from the general description in table 2.2, in terms of healthcare. Therefore, we concluded that they don't require that much of our further attention or detailed investigation.

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.

Identification: 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 shows consumers require and can attain more access rights to other consumers' or institutions' financial data, compared to patients in the healthcare use case.

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].

Scalability: Scalability is fundamental for both healthcare and finance. Nevertheless, it is important to mention that finance is far bigger industry than healthcare (as it includes healthcare itself), which means it's responsible for the well-being of far more consumers, institutions, operations and services. Easily scalable systems can bring efficiency gains and lower entry barriers for consumers. However, with Big Tech entering the financial field [Boi+21], a potential DP approach should consider measures against discrimination, abuse of market power, anti-competitive and monopolistic use of data.

Trust: It has long been recognized that trust is a key ingredient in fostering financial transaction and achieving business success. According to Nobel prize Kenneth Arrow, 'it can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence'. Since then plenty of evidence has shown that aggregate trust and aggregate economic performance are linked by a strong positive relationship. In addition, in high trust countries, institutions can grow larger [La +96] and stock markets and financial markets can prosper [GSZ08]. 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].

Confidentiality: 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].

Logging, Policies; system **Usability** and **Interoperability**; user **Ownership** of data; data **Integrity, Completeness** and **Granularity** are, of course, of great importance and are requirements that should be fulfilled in every DP approach. However, these requirements don't seem to have any specific aspects that differ from the general description in table 2.2, in terms of finance. Therefore, we concluded that they don't require that much of our further attention or detailed investigation.

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]. We differentiate 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

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

public	***private***
Ethereum [, ETH]	Hyperledger [DMH17]

Public: In public DLT designs, 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 are usually maintained by a large number of nodes, for example, Ethereum. 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: In contrast, private DLT designs 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 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+17].

	public	***private***
permissioned	-	Hyperledger [DMH17]
permissionless	Ethereum [, ETH]	-

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].

3.2 Characteristics and 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
Flexibility	⋮	The degrees of freedom in deploying applications on and customizing a distributed ledger
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
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
Security	⋮	The likelihood that functioning of the distributed ledger and stored data will not be compromised

Out of those 40 DLT characteristics, we identified 27 which correspond to our DP approach requirements described in table 2.2. This means that DLT might be suitable for DP and a potential solution some of its present issues.

Characteristic	Description
Interoperability	The ability to interact between distributed ledgers and with other external data services
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
Transaction Payload	The size of the payload in a transaction
Traceability	The extent to which transaction payloads (e.g., assets) can be traced chronologically in a DLT design
Transaction Content Visibility	The ability to view the content of a transaction in a DLT design
User Unidentifiability	The difficulty of mapping senders and recipients in transactions to identities
Node Controller Verification	The extent to which the identity of validating node controllers is verified prior to joining a distributed ledger

Auditability	The degree to which an independent third party (e.g., state institution, certification authority) can assess the functionality of a distributed ledger
Compliance	The alignment of a distributed ledger and its operation with policy requirements (e.g., regulations or industry standards)
Confirmation Latency	The time span between the inclusion of a transaction in a ledger and the point in time where enough subsequent transactions have been included in the ledger so that the likelihood of future manipulations of the initial transaction becomes negligible
Propagation Delay	The time between the submission of a transaction (or block) and its propagation to all nodes
Scalability	The capability of a distributed ledger to efficiently handle decreasing or increasing amounts of required resources (e.g., of transactions per second or number of validating nodes)
Throughput	The maximum number of transactions that can be appended to a distributed ledger in a given time interval
Transaction Validation Latency	The time required for validating a transaction by validating nodes
Ease of Node Setup	The ease of configuring and adding a new (or previously crashed) node to the distributed ledger
Ease of Use	The simplicity of accessing and working with a distributed ledger
Support for Constrained Devices	The extent to which devices with limited computing capacities can participate in a distributed ledger
Authenticity	The degree to which the correctness of data that is stored on a distributed ledger can be verified
Availability	The probability that a distributed ledger is operating correctly at any point in time
Censorship Resistance	The probability that a transaction in a distributed ledger will be intentionally aborted by a third party or processed with malicious modifications
Confidentiality	The degree to which unauthorized access to data is prevented
Consistency	The absence of contradictions across the states of the ledger maintained by all nodes participating in the distributed ledger
Durability	The property that data committed to the distributed ledger will not be lost
Integrity	The degree to which transactions in the distributed ledger are protected against unauthorized (or unintended) modification or deletion
Non-Repudiation	The difficulty of denying participation in transactions
Strength of Cryptography	The difficulty of breaking the cryptographic algorithms used in the DLT design

3.3 DLT and Data Provenance

The 27 DLT characteristics that we identified as useful for DP are very concrete and detailed. Our described requirements in table 2.2. can contain one or more DLT characteristics. Here we aim to present in an easily visible way, which DLT characteristics seem suitable for which DP requirement.

Group	Requirement	*Characteristics*
User	Identification	User Unidentifiability, Traceability, Non-Repudiation,
	Ownership	Durability,
	Accessibility	Maintainability, Node Control Verification, Transaction Content Visibility, Auditability
Data	Traceability	Traceability, Durability, Non-Repudiation, Turing-complete Smart Contracts
	Completeness	
	Granularity	Support for Constrained Devices
System	Scalability	Scalability, Maintaintability, Availability, Transaction Payload, Ease of Use, Ease of Node Setup, Throughput,
	Interoperability	Interoperability, Node Controller Verification, Support for Contrained Devices,
	Trust	Traceability, Auditability, Compliance, Authenticity, Consistancy, Durability, Non-Repudiation, Censorship Resistance, Strengt of Cryptography, Transaction Content Visibility, Turing-complete Smart Contracts
Security	Confidentiality	Confidentiality, Auditability, Transaction Content Visibility, Turing-complete Smart Contracts,
	Integrity	Integrity, Authenticity, Confirmation Latency,
	Availability	Availability, Durability, Maintainability, Throughput, Transaction Validation Latency, Propagation Delay
Other	Policies	Turing-complete Smart Contracts, Compliance,
	Usability	Ease of Node Setup, Ease of Use,
	Logging	Durability

3.4 Hyperledger Fabric

3.4.1 Features

Features:
Private
Permissioned
Scalable
Trustworthy
Transparency
Accountability
Security
Usability
Robustness
Performace
Private Transactions
Confidential Contracts
Adapted for Industry
Designed for enterprise use
Modular architecture
Pluggable consensus
Flexibility
Low Latency
Data Isolation
Multi-language smart contract support
Support for EVM and Solidity
Governance and versioning of smart contracts
Flexible endorsement model
Queryable data
Community
...

3.4.2 DLT in Healthcare

...

3.5 Ethereum

3.5.1 Features

Features:
- Public
- Permissionless
- Proof-of-Work
- Proof-of-Stake
- Transparent
- Distributed consensus among the participating entities
- Precision accuracy of the data
- Immutability and irreversibility of the ledger
- No trusted third parties necessary
- Smart Contracts
- EVM
- Immutable computer Logic
- Token (Ether) to store, process, update data via transactions
- Decentralized applications (Dapps)
- Decentralized autonomous organizations (DAOs)
- …

3.5.2 DLT in Finance

…

3.6 DLT Trade-Offs

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
	Transaction Validation Speed (+)	A	A
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*

Group	**Requirement**	***Hyperledger Fabric***	***Ethereum***
User	Identification		x
	Ownership	x	
	Accessibility	x	
Data	Traceability	x	x
	Completeness		
	Granularity		
System	Scalability	x	x
	Interoperability	x	
	Trust	x	x
Security	Confidentiality	x	
	Integrity		x
	Availability	x	x
Other	Policies	x	x
	Usability	x	x
	Logging		

4 Evaluated Mapping

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

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

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

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

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

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

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

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

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

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

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

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

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

[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).

[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.

[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.

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

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

[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.

[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.

[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.

[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).

[GGM12] Paul Groth, Yolanda Gil, and Sara Magliacane. "Automatic Metadata Annotation through Reconstructing Provenance." In: *SWPM@ ESWC*. 2012.

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

[GSZ08] Luigi Guiso, Paola Sapienza, and Luigi Zingales. "Trusting the stock market". In: *the Journal of Finance* 63.6 (2008), pp. 2557–2600.

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

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

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

[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.

[HWA10] Mohammad R Huq, Andreas Wombacher, and Peter MG Apers. "Facilitating fine grained data provenance using temporal data model". In: *Proceedings of the Seventh International Workshop on Data Management for Sensor Networks*. 2010, pp. 8–13.

[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.

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

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

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

[La +96] Rafael La Porta et al. *Trust in large organizations*. 1996.

[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.

[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.

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

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

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

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

[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.

[MZX16] Shiqing Ma, Xiangyu Zhang, and Dongyan Xu. “Protracer: Towards Practical Provenance Tracing by Alternating Between Logging and Tainting.” In: *NDSS*. 2016.

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

[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.

[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.

[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.

[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.

[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.

[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.

[Sue+13] Chun Hui Suen et al. "S2logger: End-to-end data tracking mechanism for cloud data provenance". In: *2013 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. IEEE. 2013, pp. 594–602.

[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.

[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.

[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.

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

[Wan+16] Ruoyu Wang et al. "Logprov: Logging events as provenance of big data analytics pipelines with trustworthiness". In: *2016 IEEE International Conference on Big Data (Big Data)*. IEEE. 2016, pp. 1402–1411.

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

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

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

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

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

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

[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.

[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`.