Karlsruhe Institute of Technology

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**

2 Data Provenance **3**
2.1 Definition 4
2.2 Requirements 4
2.3 Use Cases 6
2.3.1 Healthcare 6
2.3.2 Finance 7

3 Distributed Ledger Technologies **8**
3.1 Designs 9
3.2 Characteristics 11
3.3 Properties 11
3.4 DLT and Data Provenance 11
3.5 DLT in Healthcare 11
3.5.1 Current State 11
3.5.2 HyperLedger Fabric 11
3.6 DLT in Finance 11
3.6.1 Current State 11
3.6.2 Ethereum 11

4 Evaluated Mapping **12**

5 Discussion **13**
5.1 Principle Findings 13
5.2 Implications for Practice 13
5.3 Implications for Research 13
5.4 Limitations and Future Work 13

6 Conclusion **14**

Bibliography **15**

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.

Group	Requirement	Description
User	Identification	Associates each Data Subject with an unique identifier, which allows identification and lays the ground for accountability [Lee+13; Sen].
	Anonymity/ Unlinkability	Allows to send, receive or access data in an anonymous or pseudonymous way. However, provenance is an example for a possible conflict between transparency, identifiability and unlinkability [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/ Transparency	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].
	Usability	Provides clear interfaces and structures that display security aspects, required data, digital traces, policies and threats in an understandable way (usage of icons, graphs, etc.). Managing security (and privacy) is not the primary task of the user [Fre+08].
	Trustworthiness	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.
	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

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].

Group	Use Case Requirements
User	A *patient* might feel that important information should be shared, but is reluctant to do so if the information is attributed to their unique identity [, Anon]. Also, analysis of medical data is useful, but should not be done in a way that may link personal medical data to a specific *patient* or expose sensitive information, which can be used maliciously [Rin97a, p. 95]. It is important that the different actors can view, store, retrieve, move, request changes/deletion or manipulate medical data based on their ownership and access rights [Ber17](e.g. *patients* checking prescriptions, *physicians* issuing/altering prescriptions, *institutions* verifying prescriptions [, Priv]).
Data	Data Management and information on what transmitting principle was used, what type of medical data, for what purpose and to whom the information was sent is essential [Mar+20, p. 4]. It is important how medical data is collected; how, when, where it is stored, for incomplete data can impact decisions and put the *patients'* health and life at risk [KLG03]. Fine-grained provenance information helps achieve highly precise anomaly detection and auditing, which can improve decision making, diagnosing and patient safety [Muh14].
System	e-Health is a field in which big volumes of medical data are produced, exchanged and analysed [Rin97a, p. 97]. Therefore, interoperability [HK21] and usage of international standards that enforces security and patient safety are essential [Mar+20, p. 2]: the quality of the *patients'* treatment should not depend on the quality of a specific software. It's not *patients* or *physicians* job to analyse complex data flows [Rin97a, p. 97]. The system should provide clear interfaces and structures that display information in an understandable way (usage of icons, graphs, etc.). Trust is also fundamental, for that the *physician*-to-*patient* relationship is jeopardised when *patients* do not trust that their personal medical data will be kept confidential, and that this information will not be utilised for purposes other than medical [KLG03].
Security	Non-repudiation is necessary [Mar+20, p. 3]. There must not be any disclosure of medical data traveling over the network to unauthorised actors [Rin97a, p. 96]. Data must be accurate and changes should be detectable, other-wise *patients'* health and life are at risk [Mar+20, p. 10]. Also, medical data and its provenance should be available and ready for immediate use, especially in cases of emergency [KLG03].

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.

Group	Use Case Requirements
User	Without ownership or access to their own information, *consumers* cannot be certain if their data is inaccurate, obsolete, or otherwise inappropriate. [Cha85] The fear of abuse alters *consumer* behaviour and anonymity can be misused by criminals [CPS96]. A balance between identification and unlinkability must be achieved. *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]
Data	Tracing leads to transparency among actors. It should be possible to trace messages, transactions, what information and how it has been collected, analysed or processed (e.g. if donation funds are utilized properly or not). (aid) Data must be complete, accurate and fine-grained, in order to achieve precise anomaly and fraud detection and not negatively impact decision making or put *consumers*, *institutions* and their money or financial data at risk [Rua+19].
System	*Institutions* generally have an interest in maintaining good relations with *consumers* and share many of the same interests and concerns [Cha85]. To ensure trust, *institutions* need efficient, interlinked and, in a way, pervasive record-keeping system (fingerprint), while still providing *consumers* with monitoribility and control. Such systems may also have to handle a large amount of transactions [Cha85]. Easily scalable system can bring efficiency gains and lower entry barriers for *consumers*, however, there should be ways to prevent discrimination, abuse of market power, anti-competitive and monopolistic use of data [Boi+21].
Security	Where there is money related information, the actors involved are a potential subject to numerous types of crime. Non-disclosure, accuracy and availability of data, as well as state-of-the art security measures are, therefore, of great importance, in order to prevent theft, fraud, money laundering [Ung+06] or terrorist related activity [Tei18].

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].

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.

Public DLT designs, however, lack speed, face concerns over scalability and energy consumption. 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. In private networks, fewer participants means less time for the network to reach a consensus and as a result, more transactions can take place.

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.

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

3.3 Properties

3.4 DLT and Data Provenance

...

3.5 DLT in Healthcare

3.5.1 Current State

...

3.5.2 HyperLedger Fabric

...

3.6 DLT in Finance

3.6.1 Current State

...

3.6.2 Ethereum

...

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).

[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.

[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.

[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.

[CPS96] Jan Camenisch, Jean-Marc Piveteau, and Markus Stadler. “An efficient fair payment system”. In: *Proceedings of the 3rd ACM Conference on Computer and Communications Security*. 1996, pp. 88–94.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[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.

[Rua+19] Pingcheng Ruan et al. "Fine-grained, secure and efficient data provenance on blockchain systems". In: *Proceedings of the VLDB Endowment* 12.9 (2019), pp. 975–988.

[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.

[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.

[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.

[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.

[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`.