Identification of the Relationships Between Data Provenance and Blockchain as a Contributing Factor for Health Information Systems

Márcio José Sembay(✉), Douglas Dyllon Jeronimo de Macedo, and Alexandre Augusto Gimenes Marquez Filho

Federal University of Santa Catarina, Florianópolis, Brazil
marcio.sembay@posgrad.ufsc.br, douglas.macedo@ufsc.br

Abstract. On the one hand, data provenance provides the history of the origin of data and updates the modification cycle on one side, on the other hand, blockchain offers features that meet these immutability requirements of a version of data. These two technologies have applications and features that, in union, contribute to generate ideal technological structures to the management of data in several organizations. The objective of this study is to present an analysis of the relations between the main applications of data provenance as well as the blockchain features and point out applications in the Health Information Systems (HIS), in addition to a literature review on the theme. In regard to that, an analysis has been made as described in this study methodology, where the following questions were answered: i) What relations are there between data provenance and blockchain? ii) Can data provenance in union with blockchain contribute to applications in the HIS? Soon after, it was possible to prove through some of the studies present in the literature the combined use of data provenance and blockchain in HIS. Based on the relationships found between data provenance and blockchain, it was possible to conclude that these relationships contribute to data management in any organization, including HIS. It was also observed that different data provenance and blockchain methods, techniques, models, and methodologies are intertwined to generate the structures of data management in HIS, especially in Personal Health Record (PHR) and Electronic Health Record (EHR).

Keywords: Data Provenance · Blockchain · Health Information Systems

1 Introduction

The information systems and the Information and Communication Technologies (ICT), provide us with ways of communication that are changing not only our way of living but also the way businesses happens in all sectors [1]. In this context, we understand that these technologies provide the construction of one of the main human health-oriented technological scenarios, as in the case of HIS, defined as “data provenance system, information and knowledge in health care” [2].

A. L. Pinto and R. Arencibia-Jorge (Eds.): DIONE 2022, LNICST 452, pp. 258–272, 2022.
https://doi.org/10.1007/978-3-031-22324-2_20

The world health sectors are characterized by the increase of the production of data related to the patients care demands [3]. This increase in data production includes hospital records, exams results, devices that are part of the Internet of Things (IoT), among other medical data [3]. We face tons of data about varied aspects of our life, especially in health. As well as in any other sector, the health organizations are producing data to a tremendous fee, which results in many advantages and challenges all in one. Technological advances have helped generate a huge amount of data, making it a complex task when ordinal actual technology is utilized, mainly when the structures of IoT devices are geared towards user-centered design [3, 4].

In this sense, the need to manage large volumes of data in HIS meets the use of allied computational strategies that contemplate the historical processing of this data, as is the case of data provenance and blockchain. The tracking of provenance is one of the main usages of blockchain and emphasizes still that blockchain can guarantee the safety of a provenance information database [5]. Data provenance is a process which aims to provide a general view of data utilized by information systems. It focuses on the origin of data, mainly identifying the data sources and the transformations they have undergone throughout time. It is related to different application scenarios, emphasizing here the health field [6]. Data provenance usage in health has been suffering an increase in researches based on the most varied types of scientific experiments. Technologies applied to this area have obtained expressive results [7]. Data provenance is the ground to medical data quality and patients' privacy [8]. When it comes to blockchain technology, to the health field, it stands out as a record technology which is trustworthy and of consensus distributed offering alternatives to the creation of inter-operable systems, audible and safe [9]. Blockchain can be applied to the health field to control the access and distribution of sensitive information, in the transparency and audibility of service provision and data interoperability, among other situations [10].

Thus, this article analyses the main applications of data provenance defined in the work of [11] with the blockchain features presented by [12] identifying the possible existing relations between these two technologies. Eventually the analysis is finalizes with a description of some of the works found in literature which verify the usage of data provenance together with blockchain applied in HIS. Regarding this, we conclude that the analysis presented in this study that identify possible existing relations between the main applications of data provenance and blockchain features, could contribute to new studies involving these two technologies, especially in the HIS field. Therefore, the objective of this article is to present the results to this analysis and promote the use of data provenance together with blockchain in the health field, including the researches involving the prospect of these the technologies in HIS, in addition to a literature review on the theme.

This article is organized in the following manner. Section 2 describes the review of literature, where HIS, data provenance and blockchain are highly regarded. In Sect. 3, the methodology utilized in the analysis is described. Section 4 presents the analysis and identification of relations between the main applications of data provenance with the features of blockchain, and then some of the studies which verify these relations applied specifically in HIS. And finally, in Sect. 5, the conclusions and future work to this study are described.

2 Literature Review

2.1 Health Information Systems

The HIS can also be highly regarded as an information system which composes the collection, the processing, the communication, and the usage of fundamental information to perfect the efficiency of health services by means of the best management in all levels of the health sector. HIS are being implemented in all aspects of healthcare, from administration to clinical decision support systems [13]. This type of system produces relevant and good quality information to support the management and planning of health programs [2, 14]. HIS has been increasingly used around the world to improve hospital efficiency, the quality of service and patients' satisfaction [15]. HIS can also be highly regarded as an information system which integrates the collection, the processing, the communication, and the use of critical information to improve the efficiency of health services by means of a better management in all health sectors. This system produces relevant information and of good quality to support the management and planning of health programs [2, 14]. The HIS fits into two main categories. The first being the recording of individual data and the second, the collection of data for decision making and information management, usually called information management health system [16].

It is important to highlight the fact that HIS allows all the information related to the patients to be computerized, thus providing better and more efficient health services [17]. In this context, HIS is defined as a computerized means of collection, storage and recovery of information about people in health fields. Which includes patients, doctors, nurses and others who collect clinical and administrative data. This process in accomplished regardless of whether in the same or different environment in local or national levels [18–20].

Regarding this, we point out some of the main existing HIS in actuality, which are: Electronic health Record (EHR) [21], Personal Health Record (PHR) [22], Learning Health System (LHS) [23], Healthcare Monitoring System (HMS) [24], Clinical Research Information System (CRIS) [25], Hospital Information System (HIS) [26]. These HIS are utilized in several countries and are responsible for many of the processes executed in relation to the health data, contributing to a much more efficient management.

2.2 Data Provenance

Data provenance is the complementary documentation of a certain data which contains the description as "how", "when", and "why" it's been obtained and by "whom" [27]. Data provenance makes it possible to guarantee the quality and veracity of data, for example, identifying the origin of errors and attributing the sources [8]. Still, data provenance can be defined as information that helps determine the historical derivation of a data product from its origin sources, being it regarded as an essential component to allow result reproducibility, sharing and reuse of knowledge by the scientific community [28]. Now about the operation of data provenance in general, what happens is: when consultations and programs are executed the data in a data bank are moved to another data bank, and thus a description of the involving relations is created [29]. It's important to point out that in a data bank research, provenance helps in the tracking process and also in the signaling of the data origin, as well as its movement between different data sources [11].

In the context of health, the tracking of health data provenance allows the patients to have complete control over the use of all secondary personal data, in other words, creating awareness about where this data is going (for instance, public health consultations, clinical trials, among other controls in the health field) [8]. The actual health systems manage the provenance through complicated methods. These methods provide safety mechanisms that guarantee the sources of data. However, they show huge deficiencies, like depending on trustworthy third parts and vulnerability to semantics interoperability issues due to the registers of different organizations [8]. Thus, to overcome these deficiencies, data provenance models are recommended by World Wide Web Consortium[1] (W3C), named as family PROV [30] for the assistance to management of data provenance in HIS. Data provenance methods, techniques, models and methodologies are associated to different computer technologies, also observed in literature to application in HIS. In this sense, the usage of data provenance, regardless of the model, provides an essential basis for evaluation and authentication of data, offering reliability and reproducibility [31].

Regarding this, we highlight that data provenance can be applied to several scenarios [32, 33] including the health field [7]. Many research efforts in the field of data provenance induced [11] the creation of a taxonomy to categorize the efforts. The work of [11] shows that the provenance systems can be built to operate in different ways, with different characteristics and operations. Thus, we assume in the article that part of the taxonomy [11] is essential for the analysis of the relationships here considered. It also contributes to applications in the context of HIS, as has been summarized by [34]: i) Data quality: the lineage can be used to estimate the data quality and reliability based on the original data and transformation [35]. it can also offer proof declaration in data derivation [36]; ii) Audit trial: provenance can be used to track the audit trail [37], determine the usage [38] and detect mistakes in data generation [39]; iii) Replication recipes: detailed provenance information allows the repetition of data derivation, help maintain actuality [37] and be a recipe to replication [40]; iv) Attribution: pedigree can establish copyright and data propriety, make citation possible [35], and determine the responsibility in the occurrence of incorrect data; and, v) Informational: the generic usage of lineage is to consult based in lineage metadata for the discovery of data. It can also be browsed to provide a context to interpret data.

It is important to highlight that a deeper comprehension of data provenance application combined with other technologies is necessary to identify new ways of exploring its full potential.

2.3 Blockchain

Blockchain is essentially a data structure distributed, also called "public records book", in which all the confirmed transactions are stored by data sets in units called blocks. Each block refers to the previous one in chronological order, thus origination a blockchain.

[1] See https://www.w3.org.

This chain gets longer gradually as new transactions are added to the book. A mechanism of asymmetric cryptography has been applied to avoid the edition of previous blocks [41]. Blockchain in an emerging technology that has revolutionized the world scene, it was first launched in 2008 by the publication of white paper "Bitcoin: A Peer-to-Peer Electronic Cash System", when this phenomenon became popular with the creation of cryptocurrency bitcoin by [42]. However, when it comes to blockchain, there's still a certain difficulty in conceptualizing it, making it possible to find several definitions that comprise only certain aspects of this technology.

In this way, three categories are delimited in relation to the evolution of the blockchain [9]: i) Blockchain 1.0: applications in cryptocurrency, like bitcoin; ii) Blockchain 2.0: applications related to contracts of all kinds that go further than money transactions. This kind of application allows for contracts related to stock, loans, mortgage, scrip and smart contracts; and, iii) Blockchain 3.0: this category encloses the applications that go beyond the ones previously mentioned. With emphasis on areas like governmental, health, science, literature, culture, and arts. From a technical point of view, blockchain technologies are used to create a shared digital record, safe and immutable which records the history of transactions between knots in public or private peer-to-peer network. When a transaction is executed, all the knots in the network need to be in concordance about the shared record. The focus of blockchain is accounting functionality decentralized of transactions to be used to register, confirm, and transfer all kinds of contracts and properties with no need of any mediator [9].

It is important to point out that, blockchain has four features [12]: i) Immutable – (permanent and tamper-proof) a blockchain is a permanent record of transactions. Once a block is added it can't be tampered, which creates a reliable transaction record; ii) Decentralized – (networked copies) a blockchain is stored in a file that can be accessed and copied by any knot in the network, which creates decentralization; iii) Consensus Driven – (trust verification) each block in a blockchain is verified independently using consensus models that have certain rules for block validation, and often use a scarce feature (like computer power) to prove that an adequate effort has been made. In bitcoin, this process is known as mining. This mechanism works regardless of a central authority or a proxy agent; and, iv) Transparent – (full transaction history) blockchain is an open file, it can be accessed by any part and audit transactions. Which creates a provenance that makes possible to track the lifespan of actives.

Regarding this, in the health context, the potential of blockchain technology can change the way the health sector shares information, the way the medical data are updated and maintained, the way a patient's medical records are shared and updated, how the population health data are assembled and analyzed and how prescribed medication is tracked and monitored through a supply chain [43]. In the health field, blockchain can be applied to the control of access and distribution of sensitive information, in transparency and audibility of service provision and in the interoperability of data, among other situations [10]. Thus, in the works of [44] e [45], the authors reason potential blockchain contributions for the health field, which are: safety in the sharing of health data, interoperability of health data at national range, tracking of the supply chain and medical devices, tracking of the medication distribution chain, monitoring of drugs prescription, observation of aggregated events (Big Data) with main applications in collective health, patients identification, data sharing for scientific research, implementation of independent structures (e.g., management and controlling of supplementary health insurance).

3 Methodology

As for its nature, this study is considered basic research, for it is a type of research uninterested in application, inserted in a disciplinary and academic atmosphere which contrasts in all analytical aspects [46]. As to the procedures, this research is characterized as a bibliographic study, which can be defined as any research that requires the gathering of information and published material. This material may include traditional resources as books, magazines, periodicals, newspapers, and reports, but can also consist of electronic media like audio, films and video records, on-line resources like sites, blogs, and bibliographical bank data [47]. It can also be characterized, as to the objectives, like exploratory study, often named pilot and viability studies, which are considered fundamental in the evaluation of viability and value of progression for the efficiency of a study. Such studies can provide vital information to support sturdier evaluations, reducing costs and minimizing possible damages to [48]. The study also possesses a qualitative approach, with regard to the comprehension of a certain aspect of social life and its methods of (in general) generate words, instead of numbers, as data to be analyzed [49]. It is also important to point out that some methods of analysis and interpretation to be carried out in this study were based on research by [50, 51].

For the analysis of the main data provenance application relations defined by [11] with the blockchain features presented by [12], the following features have been considered: i) Highly relevant: considering the relations observed by [11] and [12] which has a direct effect on data – represented by the symbol ◎; ii) Relevant: considering the relations defined by [11] and [12] which have an indirect effect on data – represented by the symbol ⊙; and, iii) Unidentified: no relation defined by [11] and [12] – represented by the symbol ○.

Also, to prove the existing relationships between data provenance and blockchain based on the studies of [11] and [12], a literature review was carried out where five studies as of 2017 were found that presented relevant contributions to the topic in question. In this sense, the choice of these five studies for sample in this article, was based on the relevance of the existing relationships, included in some way, directly or indirectly, in the studies of [11] and [12]. Thus, for such analyzes the following questions were answered: i) What relations are there between data provenance and blockchain? and, ii) Can data provenance in union with blockchain contribute to applications in the HIS?

These relations were discovered and described in Sect. 4, considering as well the works found in literature which showed the use of data provenance and blockchain combined for the success in HIS applications.

4 Analysis

The analysis that guides this study have been accomplished in this section with the objective of identifying the relations between data provenance and blockchain and also verify through literature the direct and indirect use of such applications in HIS.

4.1 Identification of Relations Between the Main Applications of Data Provenance and Blockchain Features

Based on the works by [11] and [12], Table 1 shows a compassion between the main applications of data provenance with blockchain features in order to verify the technological harmony they both share, identifying their relations. The relations are presented below.

Table 1. Identification of the relations between data provenance and blockchain

		Blockchain characteristics			
		Transparent	Consensus Driven	Decentralized	Immutable
Main applications of Data Provenance	Informational	⊙	⊙	⊙	⊙
	Attribution	⊙	⊙	◎	◎
	Replication recipes	◎	⊙	◎	◎
	Audit trail	◎	◎	◎	○
	Data quality	⊙	⊙	⊙	⊙

◎ Highly relevant: considering the relations observed by [11] and [12] which has a direct effect on data; ⊙ Relevant: considering the relations defined by [11] and [12] which have an indirect effect on data; and, ○ Unidentified: no relation defined by [11] and [12].

As we can observe on Table 1, applications related to informational, identifies relevant relations with all features (transparent, consensus driven, decentralized and immutable). Considering the fact that this application regards data discovery in its purpose, we observe that all blockchain features are relevant when it comes to this application. For application regarding attribution, we observe a relevant relation with the blockchain features (transparent, consensus driven). Considering that this application can establish the author copyright and propriety of data, we observe that these features may contribute to a complete and verified history between data. A highly relevant relation has been identified with the features (decentralized and immutable) for this application, considering that the copied in the network and the manipulation of data must take into account the responsibility in the occurrence of erroneous data. When it comes to replication recipes with the feature (consensus driven), the identified relation is relevant, seeing as the replication of data provenance takes into account trust verification, thus allowing the guarantee of data replication for new experiments. A highly relevant relation has been identified with the features (transparent, decentralized, and immutable) for this application, considering that replication of data provenance depends on blocks which compose an immutable history record of each transaction since its origin date in a transparent and decentralized data storage. For the audit trial application, a highly relevant relation has been observed with the features (transparent, consensus driven and decentralized), considering that provenance may operate as a means of audit for data and production process, the guarantee of data audit depends on their transparency and reliability in the most diverse computer networks. There hasn't been any relation observed with the feature (immutable). Regarding application of data quality, relevant relations have been identified with all features (transparent, consensus driven, decentralized and immutable), considering that provenance can be used to obtain data quality, these data must obey certain important requirements of data provenance related to transparency and immutability of a version of data in reliable and decentralized sources. It has been observed in the analysis that audit trial applications and recipes replications both stand out relevantly with most of the features. The concentration of these applications is shown by the reasons data provenance can contribute to the guarantee, safety, confidentiality, integrity, and quality of data in several areas of knowledge.

4.2 Some of the Studies Which Verify the Relations Between Data Provenance and Blockchain Applied in HIS

In this subsection some of the studies are described that ensure and contemplate somehow the relations found in Table 1 for the health field, specifically applied to HIS.

Thus, we begin highlighting the first study, entitled "*Integrating blockchain for data sharing and collaboration in mobile healthcare applications*" [52]. The authors proposed an innovative user-centric health data sharing solution based on mobile user-controlled blockchain for cloud data sharing in a PHR. They used algorithm-based provenance data collection techniques using blockchain technologies. As a result, their solution featured an algorithm capable of managing mHealth provenance data while preserving integrity and privacy. In this study, comparing with the existing relationships in Table 1, the applications of data provenance (audit trail and data quality) with the characteristics of the blockchain (transparent, decentralized, and immutable) were the relationships that, in fact, were presented in this study. The second study is entitled "*Using PROV and Blockchain to Achieve Health Data Provenance*" [53]. The authors propose to solve the problem of managing health care data in EHR systems based on decentralized blockchain technologies and the W3C PROV model. As a result, the proposal presented by the authors was based on existing international standards ensuring interoperability between health data. The use of open systems and blockchain technology combined with the W3C PROV model together promoted the security and immutability of health records. Comparing with the existing relationships in Table 1, the applications of data provenance (informational, replication recipes, audit trail, and data quality) with blockchain characteristics (consensus driven, decentralized, and immutable) were the relationships present in this study. For the third study entitled "*Research on Personal Health data provenance and Right Confirmation with Smart Contract*" [54], to solve the leakage phenomenon, abuse, and illegal acquisition of personal health data, the authors propose a data provenance model named PROV-Chain based on blockchain technologies and Open Provenance Model (OPM). It was introduced seeking PHR applications for an IoT scenario device. As a result, the model's assessments not only guarantee the traceability of personal health data but also identified their right to personal use, guaranteeing security. In this study, comparing with the existing relationships in Table 1, the applications of data provenance (informational and attribution) with the characteristics of the blockchain (transparent, consensus driven, decentralized, and immutable) were the ones that stood out. In the fourth study entitled "*Secure and Provenance Enhanced Internet of Health Things Framework: A Blockchain Managed Federated Learning Approach*" [55], takes place in a scenario of Internet of Health Things (IoHT). For IoHT data to be acceptable to stakeholders, applications that embody IoHT must contain data accuracy, security, integrity and quality. In this sense, the authors present a hybrid federated learning model in which the intelligent blockchain contracts manage the training plan. To support the complete privacy and anonymity of IoHT data, the model has been tested with several learning applications designed for clinical trials with COVID-19[2] patients. As a result, the model in the use of applications demonstrated a strong potential for broader adoption of IoHT-based PHR health management. Comparing with the existing relationships in Table 1, the applications of data provenance (informational, audit trail, and data quality) with blockchain characteristics (consensus driven, decentralized, and immutable) were the relationships highlighted in this study. In the fifth study entitled "*Decentralised provenance for healthcare data*" [8] presents a platform for managing the provenance EHR, which can be implemented in already functioning EHR systems. The authors use blockchain technology as a basis, in addition to the Fast Healthcare Interoperability Resources (FHIR) to represent EHRs. A proxy transparently intercepts the EHRs' modifications and then triggers a smart contract[3] to perform provenance annotations using the W3C PROV language. The resulting PROV documents stored in a Hyperledger Fabric blockchain[4]. In this study, comparing with the existing relationships in Table 1, the applications of data provenance (informational, replication recipes, audit trail, and data quality) with blockchain characteristics (transparent, consensus driven, decentralized, and immutable) are highlighted.

Thus, Table 2 summarizes the studies described in this subsection, referring to the combined use of data provenance and blockchain applied in HIS.

As observed in Table 2, the technologies concerning data provenance are divided between the models (W3C PROV and OPM) and techniques of algorithm and data provenance applied to HIS. W3C PROV defines a model of data provenance to support the inter operable interchange of provenance in heterogeneous environments like the network. The PROV structure is based on the definitions of entities, activity and agents involved in data production or of something and how they are related. Thus, defining the four following types of property: *wasGeneratedBy, wasAssociatedBy, wasAttributedTo* and *used* [30]. Now considering OPM, it defines as model of open provenance that aims to characterize provenance in any "thing", material or immaterial. It seeks to show the causal relationship between events that affect objects (digital or non-digital) and describes this relation using an acyclic info graph oriented [56, 57]. It is important to highlight that the model OPM has already been overlaid by PROV. When it comes to blockchain technologies, the usage of applications is led by smart contract, followed by Hyperledger Fabric Blockchain, decentralized blockchain and health data sharing based

[2] See https://www.who.int/health-topics/coronavirus#tab=tab_1.
[3] See https://ethereum.org/en/developers/docs/smart-contracts/.
[4] See https://www.hyperledger.org/use/fabric.

Table 2. Summary of studies that combine data provenance and blockchain in HIS.

Ref.	Authors /Years	Data Provenance Technologies	Blockchain Technologies	Types of HIS	Identification of the relationships between Data Provenance and Blockchain
[8]	Margheri et al. (2020)	W3C PROV	Smart Contract and Hyperledger Fabric blockchain	EHR	Data Provenance (informational, replication recipes, audit trail, and data quality) with Blockchain (transparent, consensus driven, decentralized, and immutable)
[55]	Rayhman et al. (2020)	Algorithms based on Data Provenance	Smart Contract	PHR	Data Provenance (informational, audit trail, and data quality) with Blockchain (consensus driven, decentralized, and immutable)
[54]	Gong, Lin and Li (2019)	PROV-Chain based on the OPM	Smart Contract	PHR	Data Provenance (informational and attribution) with Blockchain (transparent, consensus driven, decentralized, and immutable)
[53]	Massi et al. (2018)	W3C PROV	Blockchain decentralized	EHR	Data Provenance (informational, replication recipes, audit trail, and data quality) with Blockchain (consensus driven, decentralized, and immutable)
[52]	Liang et al. (2017)	Algorithms based on Data Provenance	Data sharing based on blockchain	PHR	Data Provenance (audit trail and data quality) with Blockchain (transparent, decentralized, and immutable)

on blockchain. Considering now the types of HIS, the applications are directed in their most part to PHR, followed by EHR. In PHR the records are often created and monitored by the patients themselves. They can be based on desktop, internet, or smartphone devices (for example, entirely located in a cellphone or portable storage device) [22]. In EHR, collection is embracing, inter institutional and longitudinal of patients' health data. It includes, therefore, data that are not only particularly relevant for medical evaluation of one's treatment, but also for the person's health in general [21]. It is important to note that in Table 2, the relationships found between data provenance and blockchain in the five studies presented here are the same as those shown in Table 1, following the same order of relevance. It is also observed that studies show the upward growth in relation to the use of data provenance and blockchain applied in HIS, in fact contributing to the management of health data. However, it is important to highlight that, the studies here shown are limited to the technologies combined between data provenance and blockchain applied in HIS comprising the focus of the article. Other studies in the literature may bring other combined technologies in addition to those presented here, also reflecting on other successful paths applied in different health scenarios, such as the studies by [58–60].

5 Conclusions and Future Work

Data provenance applications combined with blockchain have become a challenge in HIS. The relations identified in Table 1 have concluded that there is a technological link between the main applications of data provenance combined with blockchain features in HIS. In fact, these relationships promote challenges that can help HIS, allowing to identify in each application the data provenance with the blockchain resources, which depend on the implementation or improvement of principles to obtain the expected success. Concerning this, the applications of data provenance combined with blockchain were highlighted and described in Table 2 through studies found in the literature, where we could verify that this approach applied specifically to HIS allows health organizations to store provenance records that are to be shared and tracked by the blockchain structure, thus avoiding the tampering of data. It causes to reduce the complexity of HIS which faces huge amounts of health data that depend on a safe and reliable management. Still in Table 2, it was possible to conclude through the studies described in literature, that the data provenance and blockchain technologies intertwine in methods, techniques, models, and methodologies aiming the generation of an ideal structure for the management of data in HIS. It was also possible to identify the relationships between data provenance and blockchain for each study presented in Table 2, based on the relationships in Table 1, which in fact proves the growth in the use of these two associated technologies in HIS. It has also been possible to conclude that, according to Table 2, the mainly highlighted HIS in the studies are: PHR and EHR. It being so because PHR is more convenient for patients and health workers that can monitor health data remotely by using mobile devices, especially when going through a pandemic, as is the case of the actual COVID-19. And EHR, for its use as standard medical record and being utilized in several countries in their respective HIS.

As a suggestion for future work, it is proposed the realization of a more detailed study, in the case of a systematic review of literature that might generate results of other data provenance and blockchain, beyond the ones described in this study. Furthermore, present also the main methods, techniques, models and methodologies applied in union with data provenance and blockchain used in different HIS to identify which actions could be taken to adapt technological structures ideal for these systems.

References

1. World Health Organization (WHO). Framework and standards for country health information systems. 2. ed. Geneva: WHO, (2008). ISBN 9789241595940. https://www.who.int/healthinfo/country_monitoring_evaluation/who-hmn-framework-standards-chi.pdf. Accessed 15 Jan 2022
2. Haux, R.: Health information systems-past, present, future. Int. J. Med. Inf. **75**, 268–281 (2006). https://doi.org/10.1016/j.ijmedinf.2005.08.002
3. Dash, S., Shakyawar, S.K., Sharma, M., Kaushik, S.: Big data in healthcare: management, analysis and future prospects. J. Big Data **6**(1), 1–25 (2019). https://doi.org/10.1186/s40537-019-0217-0
4. Samuel, A.M., Garcia-Constantino, M.: User-centred prototype to support wellbeing and isolation of software developers using smartwatches. Adv. Notes Inf. Sci. **1**, 140–151 (2022). https://doi.org/10.47909/anis.978-9916-9760-0-5.125
5. Greenspan, G.: Four genuine blockchain use cases. Technical report (2016). https://www.multichain.com/blog/2016/05/
6. Sembay, M.J., Macedo, D.D., Dutra, M.L.: A method for collecting provenance data: a case study in a Brazilian hemotherapy center. In: Proceedings of the 1st EAI International Conference on Data and Information in Online Environments, DIONE 2020, Florianopolis, Brazil, pp. 1–14 (2020)
7. Sembay, M.J., de Macedo, D.D.J., Dutra, M.L.: A proposed approach for provenance data gathering. Mob. Netw. Appl. **26**(1), 304–318 (2020). https://doi.org/10.1007/s11036-020-01648-7
8. Margheri, A., Massi, M., Miladi, A., Sassone, V., Rosenzweig, A.J.: Decentralised provenance for healthcare data. Int. J. Med. Inf. **141**, 1–21 (2020). https://doi.org/10.1016/j.ijmedinf.2020.104197
9. Swan, M.: Blockchain: Blueprint for a New Economy. O'Reilly Media, Newton (2015)
10. Monteil, C.: Blockchain and health. In: Digital Medicine, pp. 41–47. Springer, Cham (2019)
11. Simmhan, Y.L., Plale, B., Gannon, D.: A survey of data provenance techniques. In: Technical Report TR-618: Computer Science Department, Indiana University (2005)
12. Sultan, K., Ruhi, U., Lakhani, R.: Conceptualizing blockchains: characteristics & applications (2018). https://arxiv.org/abs/1806.03693
13. Sembay, M.J., Macedo, D.D.J.: Sistemas de informação em saúde: proposta de um método de gerenciamento de dados de proveniência no instanciamento do modelo W3C PROV-DM. Adv. Notes Inf. Sci. v. 2, Tallinn, Estonia: ColNes Publishing (2022). https://doi.org/10.47909/anis.978-9916-9760-3-6.101
14. World Health Organization (WHO). Developing Health Management Information Systems: a practical for developing countries. Manila: Regional Office for the Western Pacific (2004). https://iris.wpro.who.int/handle/10665.1/5498. Accessed 20 Nov 2021
15. Cesnik, B., Kidd, M.R.: History of health informatics: a global perspective. Stud. Health Technol. Inf. **151**, 3–8 (2010). https://doi.org/10.3233/978-1-60750-476-4-3
16. Dehnavieh, R., et al.: The district health information system (DHIS2): a literature review and meta-synthesis of its strengths and operational challenges based on the experiences of 11 countries. Health Inf. Manag. J. **48**(2) (2018). https://doi.org/10.1177/1833358318777713
17. Al Jarullah, A., El-Masri, S.: Proposal of an architecture for the national integration of electronic health records: a semi-centralized approach. Stud. Health Technol. Inf. **180**, 917–921 (2012). https://doi.org/10.3233/978-1-61499-101-4-917
18. Sligo, J., Gauld, R., Roberts, V., Villac, L.: A literature review for large-scale health information system project planning, implementation and evaluation. Int. J. Med. Inf. **97**, 86–97 (2017). https://doi.org/10.1016/j.ijmedinf.2016.09.007

19. Andargolia, A.E., Scheepers, H., Rajendran, D., Sohal, A.: Health information systems evaluation frameworks: a systematic review. Int. J. Med. Inf. **97**, 195–209 (2017). https://doi.org/10.1016/j.ijmedinf.2016.10.008
20. Robertson, A., et al.: Implementation and adoption of nationwide electronic health records in secondary care in England: qualitative analysis of interim results from a prospective national evaluation. BMJ **341**, c4564 (2010). https://doi.org/10.1136/bmj.c4564
21. Hoerbst, A., Ammenwerth, E.: Electronic health records. Meth. Inf. Med. **49**(04), 320–336 (2010). https://doi.org/10.3414/me10-01-0038
22. Liu, L.S., Shih, P.C, Hayes, G. Barriers to the adoption and use of personal health record systems. In: Proceedings of the iConference, pp. 363–370 (2011)
23. Friedman, C., et al.: Toward a science of learning systems: a research agenda for the high-functioning learning health system. J. Am. Med. Inform. Assoc. **22**(1), 43–50 (2014). https://doi.org/10.1136/amiajnl-2014-002977
24. Korhonen, I., Pärkkä, J., van Gils, M.: Health monitoring in the home of the future. IEEE Eng. Med. Biol. Magazine **22**(3), 66–73 (2003)
25. Nadkarni, P.M., Marenco, L.N., Brandt, C.A.: Clinical research information systems. Health Inf., 135–154 (2012). https://doi.org/10.1007/978-1-84882-448-5_8
26. Ismail, A., et al.: The implementation of Hospital Information System (HIS) in tertiary hospitals in Malaysia: a qualitative study Malays. J. Public Health Med. **10**(2), 16–24 (2010)
27. Buneman, P., Khanna, S., Wang-Chiew, T.: Why and where: a characterization of data provenance. In: Van den Bussche, J., Vianu, V. (eds.) ICDT 2001. LNCS, vol. 1973, pp. 316–330. Springer, Heidelberg (2001). https://doi.org/10.1007/3-540-44503-X_20
28. Freire, J., et al.: Provenance for computational tasks: a survey. J. Comput. Sci. Eng. **10**(3), 11–21 (2008). ISSN 15219615. Accessed 24 Dec 2021
29. Tan, W.C.: Provenance in databases: past, current and future. IEEE Data Eng. Bull **30**(4), 3–12 (2008). Accessed 24 May 2018
30. Gil, Y., Miles, S.: PROV Model Primer: W3C Working Draft 30 April 2013 (2013). http://www.w3.org/TR/prov-primer/
31. Moreau L., Growth, P.: Provenance: an introduction to PROV (Synthesis lectures on the semantic web: theory and technology) **3**(4), 1129 (2013). https://doi.org/10.2200/S00528ED1V01Y201308WBE007
32. Pearson, D.: Presentation on grid data requirements scoping metadata & provenance. In: Workshop on Data Derivation and Provenance, Chicago (2002)
33. Cameron, G.: Provenance and pragmatics. In: Workshop on Data Provenance and Annotation, Edinburgh (2003)
34. Goble, C.: Position statement: musings on provenance, workflow and (Semantic Web) annotations for bioinformatics. In: Workshop on Data Derivation and Provenance, Chicago (2002)
35. Jagadish, H.V., Olken, F.: Database management for life sciences research. SIGMOD Rec. **33**(2), 15–20 (2004). https://doi.org/10.1145/1024694.1024697
36. Silva, P.P. da, Silva, D., McGuinness, D.L., McCool, R.: Knowledge provenance infrastructure. In: IEEE Data Engineering Bulletin, vol. 26, pp. 26–32 (2003)
37. Miles, S., Groth, P., Branco, M., Moreau, L.: The requirements of recording and using provenance in eScience experiments. In: Technical Report, Electronics and Computer Science, University of Southampton (2005)
38. Greenwood, M., et al.: Provenance of e-science experiments: experience from bioinformatics. In: Proceedings of the UK OST e-Science second All Hands Meeting (2003)
39. Galhardas, H., Florescu, D., Shasha, D., Simon, E., Saita, C.A.: Improving data cleaning quality using a data lineage facility. DMDW, pp. 1–13 (2001). https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.22.8651&rep=rep1&type=pdf

40. Foster, I.T., Vöckler, I.S., Wilde, M., Zhao, Y.: The virtual data grid: a new model and architecture for data-intensive collaboration. In: 15th International Conference on Scientific and Statistical Database Management (2003)
41. Tian, F.: An agri-food supply chain traceability system for China based on RFID & blockchain technology. In: International Conference on IEEE Service Systems and Service Management (ICSSSM), vol. 13, pp. 1–6 (2016)
42. Nakamoto, S., et al.: Bitcoin: a peer-to-peer electronic cash system. In: Disruptive Civil Technologies: Six Technologies with Potential Impacts on Us Interests Out to 2025. NIC (National Intelligence Council) (2008). https://fas.org/irp/nic/disruptive.pdf. Accessed 28 Jan 2020
43. EY. Blockchain in health: How distributed ledgers can improve provider data management and support interoperability, p. 12 (2016). https://www.hyperledger.org/wp-content/uploads/2016/10/ey-blockchain-in-health.pdf. Accessed 10 Jan 2022
44. Bell, L., et al.: Applications of blockchain within healthcare. Blockchain Healthc. Today (2018). https://doi.org/10.30953/bhty.v1.8. Disponível em. https://blockchainhealthcaretoday.com/index.php/journal/article/view/8/40. Accessed 25 Oct 2019
45. Zhang, P., et al.: Blockchain technology use cases in healthcare. Adv. Comput. (2017). https://doi.org/10.1016/bs.adcom.2018.03.006
46. Schauz, D.: What is basic research? Insights Hist. Semant. Minerva **52**(3), 273–328 (2014)
47. Allen, M.: Bibliographic research. In: The SAGE Encyclopedia of Communication Research Methods. https://doi.org/10.4135/9781483381411.n37(2017)
48. Hallingberg, B., Turley, R., Segrott, J., et al.: Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: a systematic review of guidance. Pilot Feasibility Stud. **4**, 104 (2018). https://doi.org/10.1186/s40814-018-0290-8
49. McCusker, K., Gunaydin, S.: Research using qualitative, quantitative or mixed methods and choice based on the research. Perfusio **30**, 537–542 (2015)
50. Coimbra, F.S., Dias, T.M.R.: Use of open data to analyze the publication of articles in scientific events. Iberoamerican J. Sci. Measur. Commun. **1**(3), 1–13 (2021). https://doi.org/10.47909/ijsmc.123
51. Gontijo, M.C.A., Hamanaka, R.Y., de Araujo, R.F.: Research data management: a bibliometric and altmetric study based on Dimensions. Iberoamerican J. Sci. Measur. Commun. **1**(3), 1–19 (2021). https://doi.org/10.47909/ijsmc.120
52. Liang, X., Zhao, J., Shetty, S., Liu, J., Li, D.: Integrating blockchain for data sharing and collaboration in mobile healthcare applications. In: IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC), Montreal, Qc, Canada, pp. 1–25 (2017)
53. Massi, M., Miladi, A., Margheri, A., Sassone, V., Rosenzweig, J.: Using PROV and Blockchain to Achieve Health Data Provenance. Technical report (2018)
54. Gong, J., Lin, S., Li, J.: Research on personal health data provenance and right confirmation with smart contract. In: IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC) (2019). https://doi.org/10.1109/iaeac47372.2019.8997930
55. Rayhman, M.A., Hossain, M.S., Islam, M.S., Alrajeh, N.A., Muhammad, G.: Secure and provenance enhanced internet of health things framework: a blockchain managed federated learning approach. IEEE **8**, 205071–205087 (2020). https://doi.org/10.1109/ACCESS.2020.3037474
56. Moreau, L., Kwasnikowska, N., Van den Bussche, J.: The foundations of the open provenance model (2009). https://eprints.soton.ac.uk/267282/1/fopm.pdf
57. Open Provenance Model (OPM) (2010). https://openprovenance.org/opm/old-index.html

58. Macedo, D.D., de Araújo, G.M., de Dutra, M.L., Dutra, S.T., Lezana, Á.G. Toward an efficient healthcare CloudIoT architecture by using a game theory approach. Concurrent Eng. **27**(3), 189–200 (2019)
59. Puel, A., Wangenheim, A.V., Meurer, M.I., de Macedo, D.D.J.: BUCOMAX: collaborative multimedia platform for real time manipulation and visualization of bucomaxillofacial diagnostic images. In: IEEE 27th International Symposium on Computer-Based Medical Systems, pp.392–395 (2014). https://doi.org/10.1109/CBMS.2014.12
60. Macedo, D.D.J., de Von Wangenheim, A., de Dantas, M.A.R.: A data storage approach for large-scale distributed medical systems. In: 2015 Ninth International Conference on Complex, Intelligent, and Software Intensive Systems, pp. 486–490 (2015). https://doi.org/10.1109/CISIS.2015.88