CHAPTER 3

DATA PROVENANCE AND BLOCKCHAIN: AN APPROACH IN THE CONTEXT OF HEALTH INFORMATION SYSTEMS

Márcio José Sembay
Department of Information Science, Federal University of Santa Catarina, Brazil.
ORCID: *https://orcid.org/0000-0002-7648-8861*
Email: *marcio.sembay@posgrad.ufsc.br*

Douglas Dyllon Jeronimo de Macedo
Department of Information Science, Federal University of Santa Catarina, Brazil.
ORCID: *https://orcid.org/0000-0002-3237-4168*
Email: *douglas.macedo@ufsc.br*

Alexandre Augusto Gimenes Marquez Filho
Integrated Telemedicine and Telehealth System of Santa Catarina, Federal University of Santa Catarina, Brazil.
ORCID: *https://orcid.org/0000-0002-2656-9479*
Email: *alexandre.agmf@gmail.com*

ABSTRACT

The integration of data provenance and blockchain, in accordance with international health standards, was demonstrated to enhance patient data management through seamless integration with health information systems (HIS). This study built upon the findings of previous research conducted by the same authors, with the objective of conducting a more comprehensive and in-depth

analysis. In terms of methodology, this research was a basic study characterized as a bibliographical and exploratory investigation with a qualitative approach. The analyses carried out, based on related work, focused on the relationships between the main applications of data provenance in conjunction with the intrinsic characteristics of blockchain technology. These aspects were examined in the context of HIS, which made it possible to identify the international data interoperability standards specifically adopted in electronic health records (EHRS) and personal health records (PHRS). The primary outcomes of this study included the identification of the relationships between the primary applications of data provenance and the characteristics of blockchain, with a particular focus on HIS. Additionally, the analysis of the literature on data provenance and blockchain technology led to the recognition of the main interoperability standards. This culminated in a reflective synthesis of the findings. A comprehensive analysis of the results, grounded in the identified fundamental elements, yielded significant insights into the integration of data provenance and blockchain technology within the HIS, particularly in the context of EHR and PHR.

KEYWORDS: data provenance, blockchain, health information systems, electronic health record, personal health record

HOW TO CITE: Sembay, M. J., Dyllon Jeronimo de Macedo, D., & Augusto Gimenes Marquez Filho, A. (2025). Data provenance and blockchain: An approach in the context of health information systems. In A. Semeler (Ed.), Artificial Intelligence and Data Science Practices in Scientific Development, *Advanced Notes in Information Science, volume 8* (pp. 73-121). Pro-Metrics: Tallinn, Estonia. DOI: 10.47909/978-9916-9331-4-5.111.

1 INTRODUCTION

The health sector has been identified as a primary beneficiary of communication through information systems (IS) and information and communication technologies (ICT). These technologies have been found to support and record actions in the health context, encompassing operational, managerial, and decision support functions (World Health Organization, 2008). Consequently, the integration of ICT by competent health professionals is hypothesized to facilitate the enhancement of national health systems (Weerakoon & Chandrasiri, 2023). In this scenario, it is understood that such technologies enable the consolidation of a technological ecosystem focused on promoting human health, playing a central role in the digital transformation of care systems and the personalization of health services, as in the case of Health Information Systems (HIS), defined as "data, information, and knowledge processing systems in healthcare environments" (Haux, 2006). The global health sector is marked by an escalating volume of data pertaining to patient care requirements (Dash et al., 2019). This augmentation in data production encompasses hospital records, examination results, devices that are part of the Internet of Things (IoT), and other medical data (Dash et al., 2019). At present, we are confronted with an immense inundation of data pertaining to a myriad of aspects of life, with a particular emphasis on the healthcare sector. As in other fields, healthcare organizations have been producing data at an accelerated pace, which brings both significant benefits and challenges. Technological advances have led to exponential data generation, which has made its management a complex task, especially when using conventional technologies. This complexity is further compounded in the context of IoT devices, whose structures are guided by user-centered design (Dash et al., 2019; Samuel & Garcia-Constantino, 2022).

In this context, the growing demand for managing large volumes of data in HIS has led to the adoption of computational strategies that enable the historical processing of this information. Examples of such strategies include data provenance and the use of blockchain. The primary application of blockchain technology is tracking provenance, as it provides robust mechanisms to ensure the integrity and security of databases associated with provenance information (Greenspan, 2016). Data provenance is

a critical process for providing a comprehensive view of the data utilized in IS, with an emphasis on identifying its origins and the transformations it has undergone over time. This approach has been applied in various computational contexts, with a particular emphasis on the health domain (Sembay et al., 2020). In recent years, there has been a notable increase in the application of data provenance in scientific research focused on health-related fields, encompassing a diverse array of experiments. The technologies employed in this domain have demonstrated substantial and encouraging outcomes (Sembay et al., 2021). In this scenario, data provenance establishes itself as a fundamental foundation for ensuring the quality of medical data, as well as for strengthening the protection of patient privacy (Margheri et al., 2020).

In the domain of healthcare, blockchain technology has emerged as a reliable and consensus-based distributed ledger solution, enabling the development of interoperable, auditable, and secure systems (Swan, 2015). In the healthcare sector, its implementation entails the management of access to and dissemination of sensitive data, the enhancement of service transparency and auditability, and the assurance of data interoperability, among other pivotal applications (Monteil, 2019). Moreover, acknowledging the identified lacuna in the extant literature concerning the integration of data provenance and blockchain in HIS, this study endeavored to address three fundamental inquiries: (1) What are the conceptual and practical relationships between data provenance and blockchain technologies? (2) To what extent can the integration of data provenance mechanisms with blockchain contribute to the effectiveness, security, and interoperability of HIS? and (3) What types of data interoperability patterns can be observed in the joint use of data provenance and blockchain in HIS? These inquiries were addressed through an analytical process encompassing a theoretical review and an evaluation of practical applications within the framework of HIS. The objective of this article is to extend the research of Sembay et al. (2022). Sembay et al. conducted a study on the combined use of data provenance technologies and models and blockchain technologies employed in HIS, specifically in electronic health record (EHR) and personal health record (PHR).

It is hypothesized that this study will facilitate a more comprehensive examination of the preceding study by Sembay et al. (2022) on the primary applications of data provenance, as

delineated in the research of Simmhan et al. (2005), in conjunction with the characteristics of the blockchain as expounded by Sultan et al. (2018). This examination aims to ascertain the potential relationships between these two technological entities. Furthermore, the analysis was expanded in relation to the related work presented by Sembay et al. (2022), which addresses the joint application of data provenance and blockchain in the context of EHR and PHR. This expansion facilitated a more profound comprehension of the subject matter and enabled the identification of the predominant health data interoperability standards that have been adopted in these studies. To complement this expansion of the analysis initially developed by Sembay et al. (2022), an analytical synthesis was drawn up that makes it possible to reflect on the relevance of the elements identified, further broadening the understanding of the integrated use of data provenance and blockchain in EHR and PHR systems. Concurrently, the extant literature on the subject was expanded, thereby providing a more comprehensive basis for understanding the topic. Subsequent to this introduction, the literature review is presented, followed by the outline of the methodological approach. The results of the study are subsequently presented, summarized, and discussed. The paper concludes with a summary of its key points and a list of references.

1.1 Literature review

The literature review examines the concepts of HIS and data interoperability, with a particular emphasis on data provenance and blockchain technology. The text undertakes an examination of the primary provenance models, their applications in the health context, and their integration with standards such as Health Level 7 (HL7) Fast Healthcare Interoperability Resource (FHIR) and World Wide Web (W3C) PROV. The role of blockchain in HIS interoperability is also presented, highlighting its characteristics and applications. Consequently, studies integrating data provenance and blockchain in HIS scenarios are presented.

1.1.1 Health information systems

Health information systems are comprehensive platforms designed to collect, process, communicate, and utilize essential health data to enhance the efficiency and effectiveness of healthcare services. These systems play a crucial role in supporting management and decision-making across all levels of the healthcare sector. Health information systems are being increasingly adopted in various domains, ranging from administrative functions to clinical decision support (Sembay & Macedo, 2022). By generating high-quality and relevant information, they contribute significantly to the planning, execution, and evaluation of health programs (Haux, 2006; World Health Organization, 2004). Health information systems has been increasingly adopted across the globe to enhance hospital efficiency, the quality of service, and patient satisfaction (Cesnik & Kidd, 2010). They can also be regarded as a system of information, integrating the collection, processing, communication, and utilization of critical information. The purpose of this integration is to improve the efficiency of health services by means of enhanced management in all health sectors. This system has been demonstrated to produce relevant information of superior quality to support the management and planning of health programs (Haux, 2006; World Health Organization, 2004). The broad categorization of HIS can be subdivided into two primary classifications: systems dedicated to the recording of individual-level health data, and systems focused on the aggregation of data for decision-making and information governance, which is colloquially referred to as health information management systems (Dehnavieh et al., 2018). It is imperative to underscore that HIS facilitate the digitalization of all patient-related information, thereby enhancing the quality and efficiency of healthcare delivery (Al Jarullah & El-Masri, 2012). In this regard, HIS are characterized as a computerized system for collecting, storing, and retrieving information concerning individuals involved in the healthcare domain—including patients, physicians, nurses, and other professionals responsible for generating clinical and administrative data. This process is executed across both local and national contexts, irrespective of whether the environments are integrated or distributed (Andargolia et al., 2017; Robertson et al., 2010; Sligo et al., 2017). Regarding this, we point out some of the main existing HIS, which are as follows:

1. Electronic health record (EHR): This refers to the concept of a comprehensive, interinstitutional, and longitudinal electronic record of patient health data. This type of record includes not only information directly related to medical assessment and treatment but also data relevant to an individual's overall health status (Hoerbst & Ammenwerth, 2010). It is imperative to acknowledge that discourse pertaining to EHRs frequently pertains to the Health Insurance Portability and Accountability Act (HIPAA), a US federal statute promulgated in 1996, initially conceived to safeguard health insurance coverage for employees and their dependents (Annas, 2003).

2. Personal health record (PHR): These are health records that are frequently created and managed by the patients themselves. They may be desktop-based, web-based, or accessible via mobile devices such as smartphones or portable storage units (Liu et al., 2011).

3. Learning health system (LHS): This is a system designed to collect, share, and utilize health data to rapidly generate knowledge and support transformative decision-making that contributes to improved health outcomes. The system's operational framework is characterized by its ability to adapt to varying demands, a capability facilitated by its integration of technology, processes, and policies (Friedman et al., 2015).

4. Healthcare monitoring system (HMS): This focuses on health monitoring through the application of wearable and environmental sensors. These sensors have been developed for the purpose of collecting health-related data in patients' or users' everyday environments (Korhonen et al., 2003).

5. Clinical research information system (CRIS): This is a software system designed to support clinical research. The primary objective of CRIS is to reduce the costs of scientific studies. CRIS integrates clinical care, research data collection, and support for hospital operations (Nadkarni et al., 2012).

6. Hospital information system (HIS): In this context, the HIS can be identified as a computerized information system installed in a hospital environment with the objective of recording patient information, thereby enabling its dissemination to all sectors of the hospital that require it. An HIS is designed to support multiple functionalities, including patient care management and hospital administration, covering six distinct purposes: patient management, department management, clinical documentation, clinical decision support, financial resource management, and healthcare manager support (Ismail et al., 2010).

7. Radiology information system (RIS): This emerged with the implementation of computers in hospitals, when it was recognized that they could be used as an aid in the field of radiology (Bakker, 1991). A RIS is a specialized software designed to facilitate the management of radiology departments. It enables the reception of interpretations and the generation of patient lists. This system has the capacity to generate historical reports from radiologists and frequently transmits the final report to the HIS (Honeyman, 1999).

8. Laboratory information system (LIS): This is defined as a set of interconnected software applications designed to manage information within a clinical analysis laboratory. These applications may address technical, operational, administrative, managerial, or a combination of these aspects, with the overarching objective being the effective management of data within the laboratory setting. It is imperative to conceptualize it as an entity independent of laboratory automation systems (LAS), with which it can establish a relationship of profound intimacy, bordering on symbiosis. However, for the purpose of ascertaining its true purpose, it is essential to disengage from these systems. Laboratory automation, in turn, can be conceptualized as a component of the LAS, a comprehensive framework encompassing the management of process activities involved in the oversight of laboratory equipment and instruments, sample control, and analytical processes (Blick, 1997).

9. Picture archiving and communication system (PACS): This system consists of interconnected subsystems that utilize computer networks for the acquisition, storage, and visualization of images and data. The complexity of these systems can range from a rudimentary integration with a modality and a visualization station, accompanied by a modest database, to a sophisticated system that oversees the management of medical images across a medium or large hospital (Zhang et al., 2003). In essence, Law and Zhou (2003) offer a concise definition of the PACS as an information technology system responsible for the transmission and storage of medical images. They assert that a PACS comprises interface components for HIS/RIS, imaging modalities such as digital imaging and communications in medicine (DICOM), storage control, and viewing stations.

These HIS are implemented in various countries and play a crucial role in managing numerous processes related to health data. By facilitating the aggregation, storage, dissemination, and examination of clinical and administrative data, these systems substantially enhance the efficacy, precision, and coherence of healthcare administration. The implementation of these tools has been demonstrated to facilitate enhanced decision-making processes, improve patient care, and promote interoperability among healthcare institutions.

1.1.2 Data interoperability in HIS

The necessity for interoperability standards between HIS is intrinsic to facilitate communication and exchange of health data, thereby establishing mechanisms for interoperability among disparate health platforms. Therefore, for HIS to fulfill their role, it is essential that they possess computational tools capable of executing and mediating the entire process of interoperability of health data. In this sense, some of the most used interoperability standards in different HIS are as follows:

1. DICOM: This was developed through a collaborative effort between the American College of Radiology (ACR) and the National Electrical Manufacturers Association (NEMA). DICOM is an object-oriented standard that defines

information objects, services, and classes of services that perform these services. Each device is equipped with a set of predetermined objects that are designed to recognize the file and facilitate access to it and the associated services. Additionally, these objects enable the negotiation process between two devices to determine which one should transfer the image. The DICOM standard has been adopted by medical equipment manufacturers and healthcare informatics systems developers as the standard for exchanging images in a digital format (Honeyman, 1999; Mildenberger et al., 2002; Oosterwijk, 2002).

2. HL7 FHIR: Achieving software interoperability in the healthcare domain is possible through the implementation of consistent standards, such as HL7, a standards development organization that facilitates the exchange, integration, sharing, and retrieval of healthcare information. In this regard, the FHIR created by HL7 is another significant standard that describes data formats and elements, as well as an application programming interface for interoperable EHR exchange. Consequently, HL7 FHIR has been established as a standard that defines resources, including content definitions, architecture, models, and paradigms for exchanging health information (HL7 International, n.d.).

3. Integration of the healthcare enterprise (IHE): It was initiated in November 1998. IHE is a high-level information model designed to facilitate adaptations to the HL7 and DICOM standards. The initial objective of IHE was to establish and promote the utilization of standards, with the aim of ensuring the compliance of equipment and IS. This initiative was designed to enhance the efficiency of daily clinical operations (Huang, 2019). Consequently, the IHE makes an effective contribution to all health professionals, who can signal the main instances that emerge daily in the range of vision of their activities. While originally specified for radiology, the current objective is to establish rules for identifying and resolving the challenges that hinder the effective and functional integration of HIS. This initiative involves collaboration with medical specialists and information technology professionals. The technical architecture of the

IHE delineates a common language, vocabulary, and model using DICOM and HL7 to complete a well-defined radiological suite and clinical transactions for specific services (Bernardini et al., 2003; Huang, 2019). The objective of the IHE is to furnish the end user with enhanced access to critical and clinical patient information stored in all systems connected to a hospital network. The overarching aim is to facilitate efficiency, prognosing and integrating functionalities between incompatible systems (Boochever, 2004).

4. Extract–transform–load (ETL): This refers to a widely used process for integrating data from multiple sources or applications, including those from different domains. Extraction, transformation, and loading constitute a data management method comprising three primary phases, with the objective of preparing data for operational or analytical use. The extracted data is typically loaded into a target database, such as a data warehouse, especially for operational analytics (Bansal, 2014). The following stages comprise the fundamental phases of the process: Extract: The initial phase of the process is defined as the extraction of data from relevant data sources. These sources may be in flat file formats such as (.csv), (.xls), and (.txt), or accessed via a RESTful client. Transform: During this stage, the extracted data are cleaned and converted to comply with the schema of the target database. Common transformation tasks include data normalization, duplicate removal, integrity constraint checks, filtering based on regular expressions, data sorting and grouping, and the application of built-in functions as needed. Load: The final phase of the process involves loading the transformed data into a data warehouse. This is typically done to support Big Data environments and large-scale data analysis (Bansal, 2014).

5. Cross enterprise document sharing (XDS): This addresses the need for the registration, distribution, and access across health enterprises of patients' clinical information (Noumeir & Renaud, 2010).

6. HL7 clinical document architecture (CDA): This is a set of guidelines that define the syntax rules and provide a

fundamental framework for implementing the semantics of a clinical document. This facilitates the electronic exchange of clinical documents (Dolin et al., 2001).

1.1.3 Data provenance

Data provenance, as defined by Buneman et al. (2001), refers to the complementary documentation associated with a specific dataset. This documentation captures information about how, when, and why the data were generated, as well as by whom. This metadata plays a crucial role in ensuring the quality, authenticity, and trustworthiness of data by enabling the identification of their origin, the detection of potential errors, and the attribution of data sources (Margheri et al., 2020). Additionally, data provenance can be defined as a set of descriptive records that trace the historical derivation of a data product from its original sources. It is widely recognized as a fundamental element for ensuring the reproducibility of results, facilitating data sharing, and promoting the reuse of knowledge within the scientific community (Freire et al., 2008). In addition to its pertinence in scientific research, data provenance has also gained significance in domains such as healthcare, finance, and artificial intelligence (AI). In these fields, transparency, traceability, and accountability are paramount for compliance, auditing, and ethical data use. A fundamental aspect of data provenance is causality, which pertains to the description of the process—along with its input data and parameters—that results in the creation of a final dataset. This component is responsible for the documentation of process dependencies, thereby facilitating both the reproduction and validation of data workflows. According to Freire et al. (2008), prospective provenance specifies the intended steps to generate a data product (e.g., processes, workflows, or scripts), while retrospective provenance captures the actual execution, including system settings, inputs, outputs, and runtime parameters. In summary, prospective provenance delineates the recommended course of action, while retrospective provenance documents the actions that have been executed. This provides a foundational framework for transparency and reproducibility.

In general terms, the operation of data provenance involves tracking the movement and transformation of data during the

execution of queries and programs. In the event of such operations, data are transferred from one database to another, and a description of the relationships and processes involved is generated (Tan, 2008). In this context, data provenance is a critical element, as it facilitates the tracking of data origins, the documentation of its trajectory across various sources, and the identification of transformations and dependencies (Simmhan et al., 2005). This tracking capability is imperative for ensuring data transparency, auditability, and reliability, particularly in complex data environments.

1.1.3.1 Data provenance: Main models

To ensure the successful provenance of data in a variety of application scenarios, the creation of models will be undertaken. Consequently, initiatives to represent provenance through informational resources in general commenced with discussions on the construction of the open provenance model (OPM) in 2006, at the first International Provenance and Annotation Workshop (IPAW) (Moreau, 2006). The proposal of OPM was to define a data model that is open from an interoperability point of view, but also with respect to the community of its contributors, reviewers, and users (Moreau et al., 2009; Open Provenance Model, 2010). The OPM model aims to illustrate the causal relationship between events that impact objects (digital or otherwise) and to elucidate this relationship through a directed acyclic graph (Moreau et al., 2009; Open Provenance Model, 2010). Consequently, researchers studying OPM, in collaboration with the W3C provenance working group, have advanced their research to a new model called PROV (Moreau et al., 2011). According to Groth and Moreau (2013), the PROV document family delineates a model, serializations, and other essential supporting definitions that facilitate the exchange of provenance information in heterogeneous environments, such as the Web. The PROV family of documents comprises four recommendations: the PROV Data Model (PROV-DM), the PROV Ontology (PROV-O), the Provenance Notation (PROV-N), and Constraints of the PROV Data Model (PROV-CONSTRAINTS) (Gil & Miles, 2013; Moreau & Groth, 2013).

1.1.3.2 Data provenance in a general health context and in HIS

The application of data origin is evident in a wide range of health scenarios, which present challenges in data treatment structures. A notable example is the study by Alvarez et al. (2006), where the application of provenance occurred in the context of organ transplant administration and distribution. The work describes the development of a service-oriented architecture using provenance in medical systems to assist in the decision-making process of an organ transplant. As delineated in Li et al. (2008), an additional initiative that functions in conjunction with health data sources is the Center for Pulmonary Immunity Modeling. This initiative was established through a collaborative effort between the University of Pittsburgh, Carnegie Mellon University, and the University of Michigan. This project entailed the conceptualization and development of a data distribution platform, DataXS, which facilitates the dissemination of experimental data, analyses, and models to participating projects. This project utilizes provenance to maintain a record of the data's provenance, rather than the methodology by which the data were processed. In a recent study, Werder et al. (2022) reported concerns about the provenance of data related to applications of AI recommendations in healthcare. In their study, the authors describe several notable examples, including the use of provenance techniques integrated into EHR systems to predict sepsis, a potentially life-threatening condition in which the body's response to an infection can result in damage to its own tissues. They also discuss the application of data auditing, a practice that can be facilitated by data provenance. This allows healthcare organizations to evaluate the data used to train AI systems and identify potential diseases. Furthermore, they explore the potential of data provenance in health services to enhance understanding of the crucial factors that influence the output of a trained algorithm, such as recommending a specific diagnosis or treatment to the relevant parties.

It is imperative to underscore that, within the health context—particularly in HIS, the tracking of health data provenance empowers patients to maintain complete autonomy over the utilization of their secondary personal data. In essence, this initiative fosters transparency by providing patients with information regarding the utilization of their data in various contexts, including public health surveys, clinical trials, and other health-related

initiatives (Margheri et al., 2020). Current health systems utilize intricate mechanisms to manage provenance, implementing security measures to ensure the authenticity of data sources. However, these approaches are not without their limitations. They are dependent on trusted third parties and are vulnerable to semantic interoperability challenges arising from heterogeneous records maintained by different organizations (Margheri et al., 2020). However, it is imperative to underscore that a multitude of methods, models, and methodologies of data provenance are associated with a diverse array of computational technologies, as delineated in extant literature, to address the particular technological imperatives of HIS. In this context, the application of data provenance—independent of the HIS—provides a fundamental framework for data assessment and verification, thereby ensuring reliability and reproducibility.

1.1.3.3 Data provenance contributing to interoperability in HIS: HL7 FHIR based on W3C PROV

In the context of data provenance in HIS, it is imperative to underscore that interoperability stands as a pivotal factor to be observed for the optimal functioning of these systems, as it remains a significant challenge that persists. In this sense, HL7 FHIR utilizes provenance as a resource, indicating clinical significance in terms of confidence in the authenticity, reliability, completeness, and lifecycle stage of health data (HL7 International, n.d.). Consequently, HL7 FHIR is predicated on the W3C PROV specification, which delineates mappings of data provenance features. The W3C PROV provides design and implementation means to share semantically interoperable provenance attributes. Moreover, prominent health organizations such as IHE and HL7 endorse the W3C PROV (Margheri et al., 2020). The W3C PROV has been established as the prevailing standard for the representation of interoperable provenance information, having been adopted by HL7 FHIR (Kohlbacher et al., 2018).

1.1.3.4 Main applications of data provenance for the context of HIS

It is important to note that the concept of data provenance can be applied to a variety of scenarios, including those in the field of health (Cameron, 2003; Pearson, 2002; Sembay et al., 2021). A

considerable body of research in the domain of data provenance has given rise to the development of a taxonomy for the categorization of these efforts, as outlined by Simmhan et al. (2005). As demonstrated in the work of Simmhan et al. (2005), provenance systems can be constructed to function in various ways, exhibiting distinct characteristics and operations. Consequently, this study operates under the assumption that a component of the taxonomy delineated by Simmhan et al. (2005) is indispensable for the examination of the relationships under consideration herein. Data provenance has been shown to have a substantial impact on applications within the context of HIS, as summarized by Goble (2002):(1) Data quality: lineage can be employed to assess data quality and reliability based on the original data and its transformations (Jagadish & Olken, 2004). Additionally, it can serve as proof of data derivation (Silva et al., 2003). (2) Audit trail: provenance enables the tracking of audit trails, determining data usage, and detecting errors in data generation (Galhardas et al., 2001; Greenwood et al., 2003; Miles et al., 2005). (3) Replication recipes: detailed provenance information facilitates the replication of data derivation processes, helps maintain data currency, and acts as a guide for reproduction (Foster et al., 2003; Miles et al., 2005). (4) Attribution: provenance or pedigree can establish copyright and data ownership, enable proper citation, and assign responsibility in cases of erroneous data (Jagadish & Olken, 2004). (5) Informational: a common use of lineage metadata is to support data discovery through queries and browsing, providing contextual information necessary for data interpretation. It is important to emphasize that a deeper understanding of data provenance applications, combined with other emerging technologies, is essential to uncover new opportunities and fully exploit their potential.

1.1.4 Blockchain

Blockchain is fundamentally a distributed data structure, frequently referred to as a "public ledger," in which all confirmed transactions are stored in data units known as blocks. Each block in the blockchain contains a reference to the previous block, arranged in chronological order. This arrangement creates a continuous chain that constitutes the blockchain. This chain grows progressively as new transactions are appended to the ledger.

To guarantee the integrity and immutability of the data, blockchain employs asymmetric cryptography, which prevents the alteration of previously recorded blocks (Tian, 2016). Blockchain is an emerging technology that has caused a paradigm shift in various fields on a global scale. The concept was introduced in 2008 with the publication of the white paper "Bitcoin: A Peer-to-Peer Electronic Cash System," which popularized the concept alongside the creation of the Bitcoin cryptocurrency (Nakamoto, 2008). Despite its growing adoption, blockchain remains a complex concept, with multiple definitions emphasizing different aspects of the technology. Swan (2015) categorizes the evolution of blockchain into three distinct phases: (1) Blockchain 1.0: focused primarily on cryptocurrency applications, such as Bitcoin; (2) Blockchain 2.0: expanded applications beyond simple currency transactions to include various types of contracts, such as those related to stocks, loans, mortgages, securities, and smart contracts; (3) Blockchain 3.0: encompasses broader applications extending into domains such as government, healthcare, science, literature, culture, and the arts.

From a technical perspective, blockchain technology facilitates the establishment of a shared, secure, and immutable digital record that chronicles the history of transactions among nodes within public or private peer-to-peer networks. In the context of a transaction, it is imperative that a consensus among all network nodes is achieved to validate and record the transaction. The fundamental purpose of blockchain technology is to establish a decentralized accounting mechanism for transactions, thereby enabling the registration, verification, and transfer of various contracts and assets without the necessity of intermediaries or centralized authorities (Swan, 2015). Beyond its initial implementation in cryptocurrencies, the decentralized and tamper-resistant characteristics of blockchain have facilitated the development of transformative applications in domains such as supply chain management, voting systems, identity verification, and secure medical record-keeping. The potential of blockchain to enhance transparency, security, and trust has led to its recognition as a foundational technology for the future digital economy.

1.1.4.1 Blockchain in a general health context and in HIS

The potential of blockchain technologies to provide a unique solution for health care is significant. The broad applicability of this technology signifies its potential for integration into diverse facets of medical devices, thereby fostering advancements in various domains of health care. The healthcare sector has seen a mounting demand for blockchain technologies, with established industry players actively exploring novel applications of blockchain to address critical needs (Deloitte, 2018). One of the hallmarks of blockchain, known as immutability, is particularly vital for the storage of health data. This technology has the capacity to safeguard health records and clinical trial results, thereby ensuring regulatory compliance. The utilization of smart contracts exemplifies the application of blockchain technology in facilitating real-time patient monitoring and medical interventions (Griggs et al., 2018). In the domain of health care, blockchain technology exhibits considerable promise in its capacity to disrupt the prevailing methodologies for the management and dissemination of information. This paradigm shift has the potential to profoundly transform existing processes, including the updating and maintenance of medical data, the sharing and synchronization of patient medical records, the assembly and analysis of population health data, and the tracking of prescribed medications throughout the supply chain (Leeming, 2019). Specifically, blockchain has the potential to control access to and distribution of sensitive health information, enhance transparency and auditability of healthcare service delivery, and improve data interoperability across different systems and organizations (Monteil, 2019).

A multitude of studies, including those by Bell et al. (2018) and Zhang et al. (2017), have delineated the pivotal prospective contributions of blockchain technology to the field of health. These objectives include the assurance of data security during health information exchange, the facilitation of nationwide interoperability of health data, and the provision of reliable tracking of medical devices and supply chains. The technology under discussion has been demonstrated to facilitate the monitoring of drug prescriptions, support the surveillance of aggregated health events (leveraging Big Data analytics), and aid in patient identification and secure data sharing for scientific research purposes. Moreover, blockchain technology has the potential to facilitate

the establishment of autonomous and transparent governance structures, such as those necessary for the management and regulation of supplementary health insurance (Bell et al., 2018; Zhang et al., 2017). In the context of HIS, blockchain technology offers innovative solutions that have the potential to enhance the functionality and security of these systems to a considerable degree. At present, EMRS are generally stored in centralized data centers, with access frequently restricted to hospital networks and healthcare providers. This restriction can limit interoperability and patient control over data (Gropper, 2016). Blockchain technology is a decentralized digital ledger that utilizes cryptography for secure and transparent data storage, facilitating comprehensive and tamper-proof patient medical history records.

This approach ensures the immutability and confidentiality of medical records while concurrently enhancing the efficiency of administrative processes. For instance, blockchain has the potential to reduce the time required to resolve insurance claims and improve efficiency in generating insurance quotes by providing transparent and verifiable transaction records. Furthermore, the secure maintenance of patients' comprehensive medical histories through blockchain technology has been demonstrated to facilitate more precise and timely medication recommendations by physicians, thereby enhancing personalized healthcare services and patient safety (Gropper, 2016; Samuel, 2016). The potential applications of blockchain technology in HIS are manifold. These applications include the validation of patient data, the management of EHRS, and the tracking of research methods to manufacture safer medicines. Ensuring proper interoperability, integrity, and privacy of patient information is paramount in all of these applications. Moreover, the implementation of blockchain technology is intended to ensure transparency and auditability in the management of patient information. Most importantly, it seeks to establish robust governance frameworks that ensure proper control, accountability, and secure handling of sensitive health data throughout its lifecycle (Engelhardt, 2017; Kho, 2018; Randall et al., 2017). These mechanisms are critical to fostering trust among patients, healthcare providers, and regulatory bodies, while facilitating compliance with legal and ethical standards.

1.1.4.2 Blockchain contributing to interoperability in HIS

To maintain patient privacy in the context of data exchange with other institutions within the health ecosystem, it is imperative to implement robust access control mechanisms, ensure the integrity of data provenance, and ensure data interoperability. The interoperability of medical data between healthcare institutions and patient portals on HIS is a promising application of blockchain technology. As this technology matures, its potential to revolutionize all aspects of health care increases, and this is becoming increasingly evident (Hasselgren et al., 2020). A plethora of challenges pertaining to the interoperability of medical data across disparate HIS have been documented in the extant literature. In certain cases, there is a necessity to interweave disparate computational technologies to facilitate the exchange of data between different HIS, either due to institutional policies or the absence of a structured framework in existing standards. Therefore, it is imperative to underscore that the predominant blockchain technology solutions for the interoperability challenges encountered in disparate HIS were examined in numerous articles within the study by Peterson et al. (2016) and Zhang et al. (2018). In these articles, the authors elucidate the interoperability achieved in HIS through the utilization of HL7 FHIR-related features. An alternative approach that was identified involved the implementation of a translator component as a gateway to the data blocks, employing a different standard for translating formats (Roehrs et al., 2017).

1.1.4.3 Blockchain main features

It is important to highlight that blockchain technology possesses four fundamental features, as outlined by Sultan et al. (2018): (1) Immutable: blockchain acts as a permanent and tamper-proof ledger of transactions. Once a block is added to the chain, it cannot be altered or deleted, thereby ensuring a reliable and verifiable transaction record. (2) Decentralized: the blockchain is stored as a distributed ledger accessible and replicated across multiple nodes in the network. This decentralized architecture eliminates reliance on a central authority, enhancing resilience and reducing single points of failure. (3) Consensus-driven: each block in the blockchain is independently verified through

consensus mechanisms that define specific rules for block validation. These mechanisms often require participants to demonstrate a resource-intensive proof of work or similar effort (as exemplified by Bitcoin mining) to confirm transactions, thereby ensuring trustworthiness without the need for intermediaries. (4) Transparent: blockchain maintains a fully transparent transaction history, which is accessible to all participants in the network. This openness facilitates auditing and creates a provenance trail that allows for comprehensive tracking of the lifecycle and ownership of assets.

1.1.5 Related works combining data provenance and blockchain in HIS applications

This section presents related works that combine data provenance and blockchain technology in HIS applications. The selection of studies was made with a focus on two criteria: relevance and alignment with the overarching theme of this research.

1. The initial study, entitled "Integrating blockchain for data sharing and collaboration in mobile healthcare applications" (Liang et al., 2017), is highlighted. The authors proposed an innovative, user-centric solution for health data sharing that leverages a mobile, user-controlled blockchain framework for cloud-based PHR sharing. Their approach employs algorithm-driven techniques for data provenance collection, utilizing blockchain technology. Consequently, the solution incorporates an algorithm capable of managing the provenance of mobile health (mHealth) data while ensuring data integrity and preserving user privacy.

2. The second study is entitled "Using PROV and blockchain to achieve health data provenance" (Massi et al., 2018). The authors propose a decentralized approach to managing healthcare data in EHR systems, grounded in blockchain technologies and the W3C PROV model, as a solution to the prevailing challenges. The solution employed by the aforementioned entities utilizes recognized international standards to ensure the interoperability of health data systems. The proposed framework integrates open systems with blockchain

and the W3C PROV model, thereby enhancing the security, traceability, and immutability of health records.

3. In the third study, titled "Research on personal health data provenance and right confirmation with smart contract," the authors proposed a data provenance model called PROV-Chain. This model was developed to address issues such as data leakage, misuse, and the unauthorized acquisition of personal health information. The PROV-Chain model is built upon blockchain technologies and the OPM (Gong et al., 2019). The model has been designed for PHR applications within the context of IoT environments, with the objective of ensuring secure data sharing and accountability. The evaluation of PROV-Chain demonstrated its effectiveness in ensuring the traceability of personal health data, while also reinforcing users' rights over their own data and enhancing the overall security and integrity of HIS.

4. The fourth study, titled "Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach," is situated within the context of the Internet of Health Things (IoHT) (Rayhman et al., 2020), where ensuring data accuracy, security, integrity, and quality is fundamental for stakeholder trust and the effective adoption of IoHT-based solutions. In response to these demands, the authors propose a hybrid federated learning model in which intelligent blockchain-based smart contracts coordinate and manage the training processes. To guarantee complete privacy and anonymity of sensitive IoHT data, the proposed model was evaluated using several machine learning applications developed for clinical trials involving patients with COVID-19. The findings of the study demonstrated that the model effectively preserves data confidentiality while maintaining performance, thereby demonstrating significant potential for the broader adoption of IoHT-based PHR systems in health management.

5. The fifth study, entitled "Decentralised provenance for healthcare data," presents a platform for managing the provenance of EHRs. This platform can be implemented in existing EHR systems (Margheri et al., 2020). The authors

utilize blockchain technology in conjunction with FHIR to represent EHRs. A proxy component transparently intercepts modifications made to EHR and subsequently triggers a smart contract responsible for generating provenance annotations using the W3C PROV standard. These annotations, meticulously structured as PROV documents, are then securely recorded and stored on a hyperledger fabric blockchain. This approach ensures tamper-resistant provenance tracking, thereby enabling transparency, traceability, and verifiability of all changes applied to health records within a decentralized and auditable environment.

Consequently, the related works presented herein will serve as a foundation for some of the analyses carried out in this article, as they are potential studies with the theme addressed here.

2 METHODOLOGY

As this article constitutes an expansion of the study by Sembay et al. (2022), the methodology applied herein follows the same premises described, with certain modifications in relation to the incorporation of new analyses for novel reflections. In terms of its nature, this study is classified as basic research, as it is not primarily concerned with immediate application, but rather is embedded in an academic and disciplinary context that focuses on theoretical understanding and analytical rigor (Schauz, 2014).

With regard to the methodological procedures, the research is identified as a bibliographic study, understood as any investigation that involves the collection and analysis of information derived from previously published materials (Allen, 2017). In terms of its objectives, this study assumes an exploratory character, a quality often associated with pilot or feasibility studies. Such studies are essential in assessing the viability and potential value of progressing with a research design or intervention (Hallingberg et al., 2018). Furthermore, the study employs a qualitative approach, which aims to comprehend the dimensions of social reality through nonnumerical data, typically producing and analyzing textual information (McCusker & Gunaydin, 2015). It is also imperative to emphasize that certain analytical and interpretative methodologies employed in this research are rooted

in the frameworks and methodologies developed by Coimbra and Dias (2021) and Gontijo et al. (2021), which serve as the foundation for the critical examination of the selected literature.

To analyze the primary data provenance application relations as defined by Simmhan et al. (2005) with the blockchain features as presented by Sultan et al. (2018), the following features have been considered: (1) Highly relevant: which has a direct effect on data—represented by the symbol ⊠, (2) Relevant: which has an indirect effect on data—represented by the symbol ⊠, and (3) Unidentified: no relation defined—represented by the symbol ⊠. A literature review was conducted to examine the existing connections between data provenance and blockchain, as initially outlined in the foundational works of Simmhan et al. (2005) and Sultan et al. (2018). A comprehensive review of the extant literature was conducted, identifying five studies published between 2017 and 2020 that contributed significantly to the thematic core of this article. The selection of these related works was guided by their alignment—either direct or indirect—with the conceptual relationships proposed in the studies of Simmhan et al. (2005) and Sultan et al. (2018). Accordingly, the analytical framework of this study was structured around the following guiding research questions: (1) What are the existing relationships between data provenance and blockchain technologies? (2) How can the integration of data provenance and blockchain contribute to applications in HIS? (3) What types of data interoperability patterns emerge from the combined use of data provenance and blockchain in HIS? These inquiries were addressed through a meticulous content analysis of the selected literature, thereby facilitating a critical evaluation of the theoretical and practical intersections between provenance, blockchain, and HIS. Consequently, the subsequent section will present analyses that demonstrate how the integration of data provenance and blockchain technology contributes to the success of HIS applications, drawing upon the extant literature on the subject.

3 RESULTS

The results of the analyses presented in this section extend the findings originally reported by Sembay et al. (2022).

3.1 Identifying the relationships between key applications of data provenance and core blockchain features

As demonstrated in Table 1, a comparison is presented between the key applications of data provenance and the core features of blockchain technology. This comparison is supported by the findings of studies by Simmhan et al. (2005) and Sultan et al. (2018). The objective of this comparison is twofold: first, to verify the technological compatibility between the two approaches and, second, to identify potential points of convergence. The relationships that were identified are outlined below.

Table 1. Identification of the relations between data provenance and blockchain. **Note**. Sembay et al. (2022).

		Core features of blockchain technology			
		Transparent	Consensus-driven	Decentralized	Immutable
Key applications of data provenance	Informational	◉	◉	◉	◉
	Attribution	◉	◉	●	●
	Replication recipes	●	◉	●	●
	Audit trail	●	●	●	○
	Data quality	◉	◉	◉	◉

● Highly relevant: which has a direct effect on data;
◉ Relevant: which has an indirect effect on data;
○ Unidentified: no relation defined.

In the study by Sembay et al. (2022), the authors conducted an analysis in Table 1 to identify relationships between the main applications of data provenance and the characteristics of blockchain. The following observations were made: applications related to the informational identity demonstrated relevant relationships with all characteristics (transparent, consensus-driven, decentralized, and immutable). As demonstrated in Table 1, informational applications exhibit relevant connections with all blockchain features (transparent, consensus-driven, decentralized, and immutable), since data discovery benefits from each of these aspects. Attribution applications are closely related to transparent and consensus-driven processes, as they facilitate the establishment of authorship and ownership through the utilization of a verifiable data history. Additionally, strong links have been identified between decentralized systems and those that are immutable, given the paramount importance of accountability in the replication of data and the potential for errors to occur. In the context of replication recipes, consensus-driven mechanisms assume paramount importance, as trust verification ensures the accurate reproduction of data for new experiments. The transparency, decentralization, and immutability of blockchain technology further enhance this process by maintaining an unalterable and distributed record of transactions. In the context of audit trail applications, the attributes of transparency, consensus-driven processes, and decentralization are of paramount importance, as they ensure the traceability and reliability of data across networks. However, a direct correlation with immutable data has not been established. In the context of data quality applications, the full suite of blockchain features is pertinent, as provenance-based quality assurance depends on transparent, immutable, and decentralized data records.

The analysis indicates that audit trail and replication applications exhibit a strong alignment with blockchain capabilities, underscoring the manner in which data provenance, when integrated with blockchain, can enhance data integrity, security, confidentiality, and reliability across multiple domains. In this sense, as indicated in the analysis conducted in the study by Sembay et al. (2022), it is evident that blockchain technology can be employed to minimize aspects related to data provenance, traceability, and data authority guarantee. Indeed, the integration of data provenance with blockchain technologies has been

demonstrated to enhance data reliability and traceability, thereby providing tamper-proof information regarding the origins and transformations of data.

3.2 Relations between data provenance and blockchain applied in HIS

In this section, the related works presented in this article are analyzed based on the study by Sembay et al. (2022). The objective of this analysis is to determine whether the works consider the relationships found in Table 1, specifically applied to HIS. Consequently, Table 2 presents related studies that explore the combined application of data provenance and blockchain technology within HIS.

Table 2. Analysis of related works that combine data provenance and blockchain in HIS. **Note**. Sembay et al. (2022).

Authors/ years	**Technologies and frameworks for data provenance**	**Blockchain-based systems**	**Different forms of HIS**	**Identifying the relationships between key applications of data provenance and core blockchain features**
Margheri et al. (2020)	W3C PROV	Smart contract/hyperledger fabric blockchain	EHR	Data provenance (informational, replication recipes, audit trail, and data quality) with blockchain (transparent, consensus-driven, decentralized, and immutable)

Authors/ years	Technologies and frameworks for data provenance	Blockchain-based systems	Different forms of HIS	Identifying the relationships between key applications of data provenance and core blockchain features
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)
Gong et al. (2019)	PROV-Chain based on the OPM	Smart contract	PHR	Data provenance (informational and attribution) with blockchain (transparent, consensus-driven, decentralized, and immutable)
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)
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)

As demonstrated in Table 2 of the study by Sembay et al. (2022), the analysis emphasizes that data provenance technologies are classified into models—specifically, W3C PROV and OPM—and algorithmic techniques based on data provenance applied to HIS. The W3C PROV model facilitates interoperable exchange of provenance information across heterogeneous environments, such as networks. Its structural definition encompasses entities, activities, and agents engaged in data production or utilization, establishing four fundamental properties: *wasGeneratedBy*, *wasAssociatedBy*, *wasAttributedTo*, and *used* (Gil & Miles, 2013). In contrast, the OPM endeavors to embody provenance for all entities, irrespective of their material or immaterial nature. It does so by elucidating the causal relationships between events that exert an influence on digital or physical objects through a directed acyclic graph (Moreau et al., 2009; Open Provenance Model, 2010). It is important to acknowledge that the OPM model has since been replaced by the W3C PROV standard. With respect to blockchain technologies, the applications are predominantly driven by smart contracts, followed by hyperledger fabric blockchain, blockchain decentralized, and data sharing based on blockchain. With respect to HIS types, the majority of applications target PHR, followed by EHR. Personal health records are frequently established and overseen by patients themselves, with accessibility occurring via desktop computers, web browsers, or mobile devices, including smartphones or portable storage devices (Liu et al., 2011). Conversely, EHRS comprise extensive, interinstitutional, and longitudinal collections of patient health data, which are essential not only for clinical treatment evaluation but also for more comprehensive health management (Hoerbst & Ammenwerth, 2010).

As indicated by Table 2, the relationships identified between data provenance and blockchain across the five reviewed studies correspond closely to those in Table 1, maintaining the same order of relevance. Moreover, the extant literature suggests a growing trend in the adoption of data provenance in conjunction with blockchain technologies within HIS, which contributes positively to health data management. However, it should be emphasized that the scope of the analyzed studies is limited to specific combinations of data provenance and blockchain technologies applied to HIS, as aligned with the article's focus. It is noteworthy that other literature may present alternative combined technologies and successful implementations in various health contexts. This

assertion is supported by studies such as Puel et al. (2014), Macedo et al. (2015, 2019), and Sembay et al. (2023). Consequently, as illustrated in Table 2 of the Sembay et al. (2022) study, it is imperative to acknowledge the potential of a combined approach involving data provenance and blockchain in HIS. This integration has the capacity to induce alterations within the ecosystem of the health sector, thereby fostering trust and enhancing efficiency, thus leading to an improvement in patient treatment. Additionally, it facilitates the secure and transparent dissemination of health information stored in the HIS, thereby enhancing the accessibility of these data to external health institutions that require them to continue patient treatment. In this manner, the blockchain provides the requisite resources to guarantee data provenance in HIS. Nevertheless, challenges may arise due to technological factors, yet these present more advantages than disadvantages.

3.3 Main standards of interoperability found between data provenance and blockchain

In this section, the related works previously described are analyzed with respect to the use of the primary data interoperability standards in HIS. Consequently, Table 3 provides a comprehensive analysis of the study's findings.

Table 3. Analysis of the main standards of interoperability found in related works that combine data provenance and blockchain in HIS. **Note**. Prepared by the authors.

Authors/Years	Types of HIS	Main standards of interoperability used
Margheri et al. (2020)	EHR	HL7 FHIR, IHE, DICOM, XDS
Rayhman et al. (2020)	PHR	ETL
Gong et al. (2019)	PHR	ETL
Massi et al. (2018)	EHR	HL7 FHIR, IHE, DICOM, XDS, CDA

Authors/Years	Types of HIS	Main standards of interoperability used
Liang et al. (2017)	PHR	ETL

In the study by Margheri et al. (2020), the authors present the importance of utilizing HL7 FHIR, IHE, DICOM, and XDS in EHR. The authors posit that these interoperability standards facilitate the formulation of strategies by policymakers and project coordinators, ensuring software sustainability and safeguarding investments, while concurrently enhancing patient data security and the quality of care provided by healthcare institutions. Additionally, the use of these standards is said to optimize the combined use of data provenance and blockchain technologies in the health services offered by the HIS. In the study by Rayhman et al. (2020), the authors report the use of mHealth devices in PHR in the context of the IoHT, based on the ETL standard. This contributes to the collection of health data from various sources of mobile devices, its transformation according to the needs of the database, and its loading into a database where the necessary correlations occur for the use of these health data by the specialist professional. In this regard, the utilization of the ETL standard in the PHR facilitates the consolidation and presentation of transaction data from a data warehouse or other health database, ensuring its perpetual availability for viewing. Consequently, this enables the efficacy of data provenance processes in conjunction with blockchain technologies, particularly in IoHT scenarios, within the context of HIS.

In the study by Gong et al. (2019), the authors explore the integration of mHealth devices within hospital settings, particularly within the context of PHRs. The utilization of ETL processes plays a pivotal role in this context, facilitating the upload of comprehensive health data from these devices. Additionally, ETL contributes to the extraction of data, the maintenance of a copy of the most recent extraction, and the subsequent transfer of these data to a secure health database. Consequently, the utilization of ETL for the processes involved in data provenance and blockchain applied in the context of PHR contributes to the tracking of health data in these scenarios. In the study by Massi et al. (2018), the authors mention the use of HL7 FHIR, IHE, DICOM, XDS, and

CDA to contribute to the existing limitations of interoperability in EHR systems. The authors posit that the established criteria contribute to the system they have proposed, which utilizes blockchain technology to manage the provenance of health documents. This system is designed to seamlessly integrate into existing EHR deployments. In the study conducted by Liang et al. (2017), the authors posited that the dissemination of health data among institutions necessitates the establishment of a secure data sharing infrastructure. However, there are several challenges related to privacy, security, and interoperability. In this regard, the utilization of the ETL standard for mHealth devices to populate health databases plays a pivotal role in facilitating the implementation of blockchain technologies for the management of data provenance in PHR and the development of novel iterations of EHR, featuring user-centric access control and privacy preservation mechanisms.

Therefore, the interoperability standards highlighted in each study analyzed in Table 3 demonstrate that they are critical requirements for HIS. Notwithstanding the extant limitations that may imperil patient safety, the interoperability standards delineated in Table 3 are the most widely utilized and contribute to ameliorating the preponderance of limitations in the exchange of health data between disparate HIS. In conclusion, an evident correlation was identified among the patterns exhibited in Table 3. These patterns collectively contributed to enhancing the requirements concerning the tracking of health data, as well as the security and immutability of these data when utilizing blockchain technologies in HIS.

3.4 Summary and reflections of the analysis presented

This section presents a synopsis of the analysis performed, as illustrated in Figure 1. The analysis was conducted using data from Tables 1–3, and it highlights the significance of the elements identified during the course of the study.

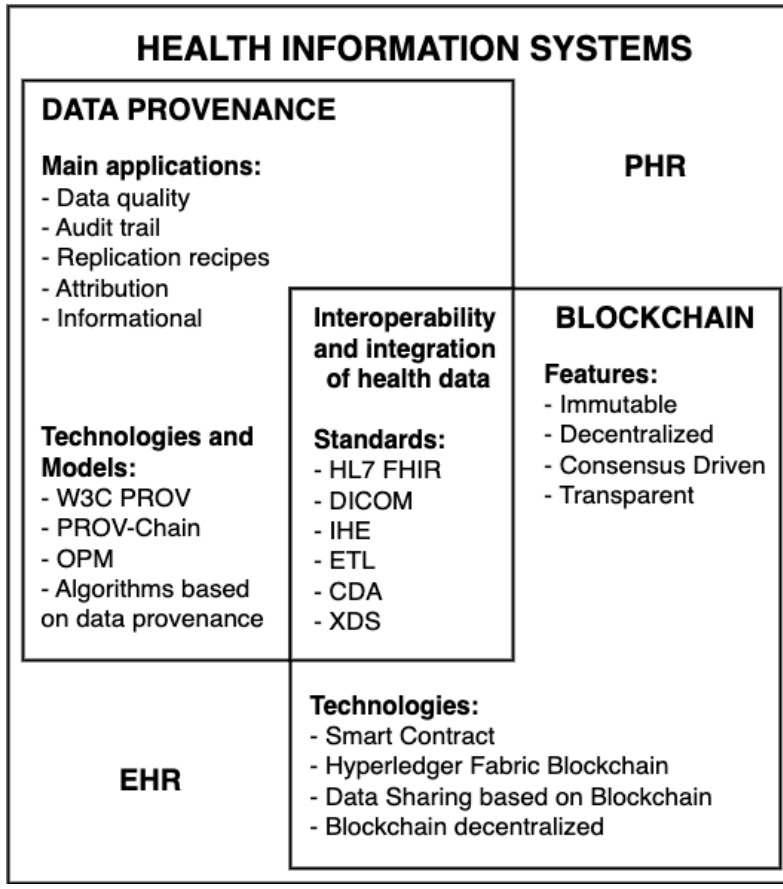

Figure 1. Summary of analysis. **Note**. Prepared by the authors.

Figure 1 presents the primary elements identified in the analysis, underscoring the significance of integrating data provenance and blockchain within the framework of HIS. With regard to Figure 1, the following observations can be made: (1) It was observed in the studies that the EHR and PHR are, in fact, the most used HIS. Making a general analysis of these two systems, the following reflections stand out: the EHR is the most used by doctors to improve the quality of care, having as its main advantage the

availability of medical information between providers; the EHR and PHR reside on different platforms under various technologies and standards; and PHR allows the integration of the main information components in the EHR systems. Thus, it is important to emphasize that the integration of medical information into EHR and PHR leads to a dramatic change in personalized care; (2) Regarding the main applications of data provenance (data quality, audit trial, replication recipes, attribution, and informational) that intertwine with blockchain characteristics (immutable, decentralized, consensus-driven, and transparent) in the context of HIS, it can be stated that data provenance and blockchain when combined in the context of HIS, mainly in EHR and PHR, result in benefits for this context. In this sense, data provenance is the foundation of medical data quality and patient privacy, and blockchain contributes to the creation and management of provenance records, both in the context of EHR and PHR; and (3) Regarding data provenance technologies and models (W3C PROV, PROV-Chain, OPM, and algorithms based on data provenance) together with blockchain technologies (smart contract, hyperledger fabric blockchain, data sharing based on blockchain, and blockchain decentralized) result in several challenges encountered in data interoperability issues in EHR and PHR. A significant challenge confronting the field is the establishment of system interoperability, defined as the standardization of data and information that can be read, understood, and accessed from any health unit, whether public or private. In this sense, the utilization of these technologies and models of data provenance and blockchain underscores the nexus that pertains to interoperability and integration of health data (HL7 FHIR, DICOM, IHE, ETL, CDA, and XDS).

While these standards do not address all interoperability issues, they aim to align with the requirements of EHR and PHR, thereby enhancing the quality of care and the efficiency of healthcare services. Furthermore, these standards are designed to facilitate the secure exchange of health data between EHRS and PHRS, thereby promoting interoperability and enhancing patient care.

4 DISCUSSION

The findings of this study underscore the significance of integrating data provenance and blockchain technology to enhance efficiency, security, and interoperability in HIS, particularly within the domains of EHR and PHR. As the analysis indicates, EHRS remain the primary instrument utilized by healthcare professionals to ensure the delivery of quality care. Conversely, PHRS promote the integration of patient information across multiple platforms, thereby fostering a more personalized care approach. The intersection of data provenance and blockchain has demonstrated considerable potential. Provenance contributes attributes such as traceability, auditing, information quality, and attribution of authorship, which are essential for guaranteeing the integrity of medical data. The blockchain technology under discussion in this paper is characterized by its immutability, decentralization, and transparency. These characteristics contribute to the establishment of a robust layer of security and reliability. The integration of these technologies, as Sembay et al. (2022) have noted, has been demonstrated to enhance the creation of reliable and auditable records, thereby reducing risks and strengthening confidence in the use of HIS. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The analysis identified that, despite the potential of these technologies, significant challenges related to interoperability persist. The heterogeneity of standards and platforms adopted by EHR and PHR systems engenders challenges in the seamless integration of data. Technologies such as W3C PROV, PROV-Chain, OPM, smart contracts, and hyperledger fabric blockchain, when associated with interoperability standards such as HL7 FHIR, DICOM, IHE, CDA, and XDS, seek to mitigate these challenges. Despite these advancements, the pursuit of complete standardization remains unfinished, as technical and institutional barriers persist. From a pragmatic standpoint, the findings of this study underscore the feasibility of enhancing health data integration through the joint implementation of data provenance and blockchain technology. This approach is proposed as a method to guarantee the integrity, security, and effective dissemination of information. The study makes a theoretical contribution

to the field by offering insights into the complementarity between these technologies. This understanding can inform the development of future models for more secure and interoperable HIS architecture. From a political standpoint, the findings underscore the necessity for public and regulatory policies that promote the utilization of open standards and reliable technologies for health data management.

However, it is important to note that this research is not without its limitations. The analysis was primarily based on secondary studies and a limited sample of related work. It should be noted that no empirical experiments or direct surveys were conducted in actual HIS environments. Furthermore, given the perpetual evolution of technology, it is important to note that the results may not fully capture the most recent advancements. These factors may limit the generalizability of the findings and necessitate caution when extrapolating the results. In summary, the findings suggest that integrating data provenance with blockchain technology holds potential for enhancing the quality, security, and interoperability of HIS. The future of this field will be determined by three factors: greater standardization, practical experimentation, and collaboration between technological agents, health professionals, and public policy makers.

5 CONCLUSION

The analysis described in this article suggests that data provenance applications combined with blockchain have the potential to be promising in a variety of application sectors, as illustrated in Table 1. In this sense, as illustrated in Table 1, it was possible to understand that the relationships found may be directed to the HIS, specifically the EHR and PHR, as shown in Table 2 in the analysis carried out in the works presented. Therefore, as indicated by the findings presented in Table 2, it is evident that alterations in the EHR and PHR ecosystem may transpire. To address this, it is imperative to identify suitable models, methods, techniques, and methodologies that will empower health organizations to store provenance records. These records, in turn, must be shared and tracked by the blockchain structure, thereby mitigating the risk of data tampering. This approach serves to mitigate the complexity encountered by HIS when confronted with

substantial volumes of health data, which necessitates secure and reliable management. The integration of blockchain technology with data provenance holds considerable promise in this regard. Furthermore, an analysis of Table 2 suggests that the most prominent HIS in the studies are: EHR and PHR. This is because the EHR is used as the standard medical record used in several countries in their respective HIS, and the PHR is the most convenient for patients and healthcare professionals who can monitor health data remotely via mobile devices, especially in times of pandemic, as was the case with COVID-19.

Another salient point pertains to the data interoperability standards delineated in Table 3, which concerns the amalgamation of data provenance and blockchain in HIS, particularly in the context of EHR and PHR. These standards contribute to the normalization and interoperability of health data in the aforementioned HIS. However, challenges persist, particularly with regard to the security and privacy of patient data. This indicates that, given the existence of multiple HIS, health institutions have prioritized standardizing clinical procedures to ensure uniformity in practice and establishing systems to facilitate the exchange of data and information across different HIS. The analysis, as depicted in Figure 1, demonstrates that the integration of data provenance and blockchain in EHR and PHR systems, despite the challenges associated with this integration, offers significant benefits. Figure 1 underscores a mounting trend in the implementation of blockchain technology for the management of healthcare document provenance, exhibiting the capacity for seamless integration across disparate healthcare institutions. This approach facilitates the secure management of document provenance while ensuring data privacy.

Finally, as a suggestion for future research, it is recommended to undertake a more comprehensive and detailed systematic literature review that extends beyond the scope of this study. A rigorous investigation should be undertaken to ascertain the existence of any additional applications and integrations of data provenance combined with blockchain technology across various HIS. This includes a thorough presentation of the principal methods, techniques, models, and methodologies employed in conjunction with data provenance and blockchain within different HIS contexts. Furthermore, subsequent research endeavors should investigate the operational dynamics of integrating data

provenance and blockchain within HIS environments that incorporate advanced technologies, such as cloud computing and the IoHT. Additionally, these studies should examine their interactions with AI applications in medicine. It is imperative to comprehend these relationships to ascertain how this technological synergy can enhance data security, interoperability, transparency, and decision-making processes in healthcare. Indeed, a more in-depth study on the combined use of data provenance and blockchain has the potential to offer significant contributions by clarifying the specific challenges faced in these implementations. Furthermore, it would facilitate the identification of the most effective technological architectures and frameworks that maximize the benefits of this integration, thereby supporting the successful deployment and adoption of these innovations across diverse HIS.

Conflict of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Contribution statement

Márcio José Sembay: Writing – Review & Editing, Writing – Original Draft.
Alexandre Augusto Gimenes Marquez Filho: Methodology, Conceptualization.
Douglas Dyllon Jeronimo de Macedo: Writing – Review & Editing, Supervision.

Statement of data consent

This study did not generate any new or large-scale datasets requiring deposition in repositories. The research is based exclusively on the analysis and interpretation of related works previously published in the literature, which supported the development and extension of this manuscript.

REFERENCES

Al Jarullah, A., & El-Masri, S. (2012). Proposal of an architecture for the national integration of electronic health records: A semi-centralized approach. *Studies in Health Technology and Informatics, 180*, 917–921. *https://doi.org/10.3233/978-1-61499-101-4-917*

Allen, M. (2017). Bibliographic research. In *The SAGE Encyclopedia of communication research methods*. SAGE Publications. *https://doi.org/10.4135/9781483381411.n37*

Alvarez, S., Vazquez-Salceda, J., Kifor, T., Varga, L. Z., & Willmott, S. (2006). Applying provenance in distributed organ transplant management. In *Provenance and annotation of data: International provenance and annotation workshop, IPAW 2006, Chicago, IL, USA, May 3–5, Revised Selected Papers* (pp. 28–36). Springer Berlin Heidelberg. *https://doi.org/10.1007/11890850_4*

Andargolia, A. E., Scheepers, H., Rajendran, D., & Sohal, A. (2017). Health information systems evaluation frameworks: A systematic review. *International Journal of Medical Informatics, 97*, 195–209. *https://doi.org/* 10.1016/j.ijmedinf.2016.10.008

Annas, G. J. (2003). HIPAA regulations: A new era of medical-record privacy? *New England Journal of Medicine, 348*(15), 1486. *https://doi.org/10.1056/NEJMlim035027*

Bakker, A. R. (1991). HIS, RIS, and PACS. *Computerized Medical Imaging and Graphics, 15*(3), 157–160. *https://doi.org/10.1016/0895-6111(91)90004-F*

Bansal, S. K. (2014). Towards a semantic extract-transform-load (ETL) framework for big data integration. In *2014 IEEE international congress on big data* (pp. 522–529). IEEE. http://10.1109/BigData.Congress.2014.82

Bell, L., Buchanan, W. J., Cameron, J., & Lo, O. (2018). Applications of blockchain within healthcare. *Blockchain in Healthcare Today, 1*, 1–7. *https://doi.org/10.30953/bhty.v1.8*

Bernardini, A., Alonzi, M., Campioni, P., Vecchioli, A., & Marano, P. (2003). IHE: Integrating the Healthcare Enterprise, towards complete integration of healthcare information systems. *Rays, 28*(1), 83–93. *https://pubmed.ncbi.nlm.nih.gov/14509182/*

Blick, K. E. (1997). Decision-making laboratory computer systems as essential tools for achievement of total quality. *Clinical Chemistry, 43*(5), 908–912. *https://doi.org/10.1093/clinchem/43.5.908*

Boochever, S. S. (2004). HIS/RIS/PACS integration: Getting to the gold standard. *Radiology Management, 26*(3), 16–24. *https://pubmed.ncbi.nlm.nih.gov/15259683/*

Buneman, P., Khanna, S., & Wang-Chiew, T. (2001). Why and where: A characterization of data provenance. In J. Van den Bussche & V. Vianu (Eds.), *ICDT 2001: International conference on database theory* (pp. 316–330). Springer. *https://doi.org/10.1007/3-540-44503-X_20*

Cameron, G. (2003). *Provenance and pragmatics* [Workshop on Data Provenance and Annotation]. Edinburgh, UK.

Cesnik, B., & Kidd, M. R. (2010). History of health informatics: A global perspective. *Studies in Health Technology and Informatics, 151*, 3–8. *https://doi.org/10.3233/978-1-60750-476-4-3*

Coimbra, F. S., & Dias, T. M. R. (2021). Use of open data to analyze the publication of articles in scientific events. *Iberoamerican Journal of Science Measurement and Communication, 1*(3), 1–13. *https://doi.org/10.47909/ijsmc.123*

Dash, S., Shakyawar, S. K., Sharma, M., & Kaushik, S. (2019). Big data in healthcare: Management, analysis and future prospects. *Journal of Big Data, 6*(1), 1–25. *https://doi.org/10.1186/s40537-019-0217-0*

Dehnavieh, R., Haghdoost, A., Khosravi, A., Hoseinabadi, F., Rahimi, H., Poursheikhali, A., Shafiee, G., Gholami, H., Abadi, M. B. H., Noori, R., & Mehrolhassani, M. H. (2018). 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 Information Management Journal, 48*(2), 62–75. *https://doi.org/10.1177/1833358318777713*

Deloitte. (2018). *Breaking blockchain open: Deloitte's 2018 global blockchain survey* (Report No. 48). Deloitte Insights. *https://doi.org/10.1002/ejoc.201200111*

Dolin, R. H., Alschuler, L., Beebe, C., Biron, P. V., Boyer, S. L., Essin, D., & Mattison, J. E. (2001). The HL7 clinical document architecture. *Journal of the American Medical Informatics Association, 8*(6), 552–569. *https://doi.org/10.1136/jamia.2001.0080552*

Engelhardt, M. A. (2017). Hitching healthcare to the chain: An introduction to blockchain technology in the healthcare sector. *Technology Innovation Management Review, 7*(10), 22–34. *http://doi.org/10.22215/timreview/1111*

Foster, I. T., Vöckler, J.-S., Wilde, M., & Zhao, Y. (2003). *The virtual data grid: A new model and architecture for data-intensive collaboration.* In *15th International conference on scientific and statistical database management (SSDBM), Cambridge, MA, USA*. *https://www.cidrdb.org/cidr2003/program/p18.pdf*

Freire, J., Silva, C. T., Callahan, S. P., Santos, E., Scheidegger, C. E., & Vo, H. T. (2008). Provenance for computational tasks: A survey. *Journal of Computing Science and Engineering, 10*(3), 11–21. *https://doi.org/10.1109/MCSE.2008.79*

Friedman, C., Rubin, J., Brown, J., Buntin, M., Corn, M., Etheredge, L., Gunter, C., Musen, M., Platt, R., Stead, W., Sullivan, K., & Van Houweling, D. (2015). Toward a science of learning systems: A research agenda for the high-functioning learning health system. *Journal of the American Medical Informatics Association, 22*(1), 43–50. *https://doi.org/10.1136/amiajnl-2014-002977*

Galhardas, H., Florescu, D., Shasha, D., Simon, E., & Saita, C. A. (2001). *Improving data cleaning quality using a data lineage facility*. In *Proceedings of the international workshop on design and management of data warehouses (DMDW), Interlaken, Switzerland* (pp. 1–13). *http://ceur-ws.org/Vol-39/paper3.pdf*

Gil, Y., & Miles, S. (2013). *PROV model primer* [W3C Working Draft]. W3C. *https://www.w3.org/TR/prov-primer/*

Goble, C. (2002). *Position statement: Musings on provenance, workflow and (Semantic Web) annotations for bioinformatics* [Workshop on Data Derivation and Provenance]. Chicago, IL, USA.

Gong, J., Lin, S., & Li, J. (2019). Research on personal health data provenance and right confirmation with smart contract. In *IEEE 4th advanced information technology, electronic and automation control conference (IAEAC). https://doi.org/10.1109/IAEAC47372.2019.8997930*

Gontijo, M. C. A., Hamanaka, R. Y., & de Araujo, R. F. (2021). Research data management: A bibliometric and altmetric study based on Dimensions. *Iberoamerican Journal of Science Measurement and Communication, 1*(3), 1–19. *https://doi.org/10.47909/ijsmc.120*

Greenspan, G. (2016). *Four genuine blockchain use cases* [Technical report]. MultiChain. *https://www.multichain.com/blog/2016/05*

Greenwood, M., Goble, C., Stevens, R., Zhao, J., Addis, M., Marvin, D., Moreau, L., & Oinn, T. (2003). *Provenance of e-science experiments: Experience from bioinformatics*. In *Proceedings of the UK OST e-science second all hands meeting, Nottingham, UK. https://eprints.soton.ac.uk/258895/1/prov-all-hands.pdf*

Griggs, K. N., Ossipova, O., Kohlios, C. P., Baccarini, A. N., Howson, E. A., & Hayajneh, T. (2018). Healthcare blockchain system using smart contracts for secure automated remote patient monitoring. *Journal of Medical Systems, 42*(7), 130. *https://doi.org/10.1007/s10916-018-0982-x*

Gropper, A. (2016). *Powering the physician-patient relationship with HIE of one blockchain health IT* [ONC/NIST Use of Blockchain for Healthcare and Research Workshop]. Gaithersburg, MD. *https://www.healthit.gov/sites/default/files/7-29-poweringthephysician-patientrelationshipwithblockchainhealthit.pdf*

Groth, P., & Moreau, L. (2013). *PROV-overview: An overview of the PROV family of documents*. W3C. *https://www.w3.org/TR/prov-overview/*

Hallingberg, B., Turley, R., Segrott, J., Wight, D., Craig, P., Moore, L., Murphy, S., Robling. M., Simpson, S. A., & Moore, G. (2018). Exploratory studies to decide whether and how to proceed with full-scale evaluations of public health interventions: A systematic review of guidance. *Pilot and Feasibility Studies, 4*, 104. *https://doi.org/10.1186/s40814-018-0290-8*

Hasselgren, A., Kralevska, K., Gligoroski, D., Pedersen, S. A., & Faxvaag, A. (2020). Blockchain in healthcare and health sciences: A scoping review. *International Journal of Medical Informatics, 134*, Article 104040. *https://doi.org/10.1016/j.ijmedinf.2019.104040*

Haux, R. (2006). Health information systems—Past, present, future. *International Journal of Medical Informatics*, *75*(3–4), 268–281. *https://doi.org/10.1016/j.ijmedinf.2005.08.002*

HL7 International. (n.d.). *Fast healthcare interoperability resources release (STU)*. *http://fhir.hl7.org*

Hoerbst, A., & Ammenwerth, E. (2010). Electronic health records. *Methods of Information in Medicine, 49*(4), 320–336. *https://doi.org/10.3414/me10-01-0038*

Honeyman, J. C. (1999). Information systems integration in radiology. *Journal of Digital Imaging, 12*(Suppl 1), 218–219. http://doi: 10.1007/BF03168810

Huang, H. K. (2019). *PACS-based multimedia imaging informatics: Basic principles and applications* (3rd ed.). John Wiley & Sons. *https://doi.org/10.2345/i0899-8205-40-2-125.1*

Ismail, A., Jamil, A. T., Rahman, A. F. A., Bakar, J. M. A., Saad, N. M., & Saadi, H. (2010). The implementation of Hospital Information System (HIS) in tertiary hospitals in Malaysia: A qualitative study. *Malaysian Journal of Public Health Medicine, 10*(2), 16–24.

Jagadish, H. V., & Olken, F. (2004). Database management for life sciences research. *ACM SIGMOD Record, 33*(2), 15–20. *https://doi.org/10.1145/1024694.1024697*

Kho, W. (2018). Blockchain revolution in healthcare: The era of patient-centred dental information system. *International Journal of Oral Biology, 43*(1), 1–3. *https://doi.org/10.11620/IJOB.2018.43.1.001*

Kohlbacher, O., Mansmann, U., Bauer, B., Kuhn, K., & Prasser, F. (2018). Data Integration for Future Medicine (DIFUTURE): An architectural and methodological overview. *Methods of Information in Medicine, 57*(S01), e43–e50. *https://doi.org/10.3414/ME17-02-0022*

Korhonen, I., Pärkkä, J., & van Gils, M. (2003). Health monitoring in the home of the future. *IEEE Engineering in Medicine and Biology Magazine, 22*(3), 66–73. *https://doi.org/10.1109/MEMB.2003.1213628*

Law, M. Y., & Zhou, Z. (2003). New direction in PACS education and training. *Computerized Medical Imaging and Graphics, 27*(2–3), 147–156. *https://doi.org/10.1016/S0895-6111(02)00088-5*

Leeming, G., Ainsworth, J., & Clifton, D. A. (2019). Blockchain in health care: Hype, trust, and digital health. *The Lancet*, 393(10190), 2476–2477. *https://doi.org/10.1016/S0140-6736(19)30948-1*

Li, Q., Labrinidis, A., & Chrysanthis, P. K. (2008). User-centric annotation management for biological data. In *Provenance and annotation of data and processes: second international provenance and annotation workshop, IPAW 2008, Salt Lake City, UT, USA, June 17–18, Revised Selected Papers* (pp. 54–61). Springer Berlin Heidelberg. *https://doi.org/10.1007/978-3-540-89965-5_7*

Liang, X., Zhao, J., Shetty, S., Liu, J., & Li, D. (2017). *Integrating blockchain for data sharing and collaboration in mobile healthcare applications.* In *2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC), Montreal, QC, Canada* (pp. 1–25). IEEE. *https://doi.org/10.1109/PIMRC.2017.8292361*

Liu, L. S., Shih, P. C., & Hayes, G. (2011). Barriers to the adoption and use of personal health record systems. *Proceedings of the iConference*, 363–370. *https://doi.org/10.1145/1940761.1940811*

Macedo, D. D. J., de Von Wangenheim, A., & de Dantas, M. A. R. (2015). 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). *https://doi.org/10.1109/CISIS.2015.88*

Macedo, D. D., de Araújo, G. M., de Dutra, M. L., Dutra, S. T., & Lezana, Á. G. (2019). Toward an efficient healthcare Cloud IoT architecture by using a game theory approach. *Concurrent Engineering, 27*(3), 189–200. *https://doi.org/10.1177/1063293X19844548*

Margheri, A., Massi, M., Miladi, A., Sassone, V., & Rosenzweig, A. J. (2020). Decentralised provenance for healthcare data. *International Journal of Medical Informatics, 141*, Article 104197. *https://doi.org/10.1016/j.ijmedinf.2020.104197*

Massi, M., Miladi, A., Margheri, A., Sassone, V., & Rosenzweig, J. (2018). *Using PROV and blockchain to achieve health data provenance* [Technical Report]. University of Southampton. *https://eprints.soton.ac.uk/421292/1/PROV_BC_Healthcare.pdf*

McCusker, K., & Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. *Perfusion, 30*(7), 537–542. *https://doi.org/10.1177/0267659114559116*

Mildenberger, P., Eichelberg, M., & Martin, E. (2002). Introduction to the DICOM standard. *European Radiology, 12*(4), 920–927. http://doi: 10.1007/s003300101100

Miles, S., Groth, P., Branco, M., & Moreau, L. (2005). *The requirements of recording and using provenance in eScience experiments* [Technical Report]. Electronics and Computer Science, University of Southampton, UK. *https://eprints.soton.ac.uk/260269/1/pasoa04requirements.pdf*

Monteil, C. (2019). Blockchain and health. In *Digital medicine* (pp. 41–47). Springer. *https://doi.org/10.1007/978-3-319-98216-8_4*

Moreau, L. (2006). *Usage of "provenance": A tower of Babel* [Position paper]. Microsoft Life Cycle Seminar, Mountain View, CA. *https://eprints.soton.ac.uk/409382/*

Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., & Van den Bussche, J. (2011). The open provenance model core specification (v1.1). *Future Generation Computer Systems, 27*(6), 743–756. *https://doi.org/10.1016/j.future.2010.07.005*

Moreau, L., & Groth, P. (2013). *Provenance: An introduction to PROV*. Morgan & Claypool. *https://doi.org/10.2200/s00528ed1v01y201308wbe007*

Moreau, L., Kwasnikowska, N., & Van den Bussche, J. (2009). *The foundations of the open provenance model*. University of Southampton. *https://eprints.soton.ac.uk/267282/1/fopm.pdf*

Nadkarni, P. M., Marenco, L., & Brandt, C. (2012). Clinical research information systems. In *Health informatics* (pp. 135–154). Springer. *https://doi.org/10.1007/978-1-84882-448-5_8*

Nakamoto, S. (2008). *Bitcoin: A peer-to-peer electronic cash system.* National Intelligence Council. *https://fas.org/irp/nic/disruptive.pdf*.

Noumeir, R., & Renaud, B. (2010). IHE cross-enterprise document sharing for imaging: Interoperability testing software. *Source Code for Biology and Medicine, 5*(1), 1–15. *https://doi.org/10.1186/1751-0473-5-9*

Oosterwijk, H. (2002). *DICOM basics* (2nd ed.). Otech.

Open Provenance Model (OPM). (2010). *Open Provenance Model (OPM) specifications*. *https://openprovenance.org/opm/old-index.html*

Pearson, D. (2002). *Presentation on grid data requirements scoping metadata & provenance* [Workshop on Data Derivation and Provenance], Chicago, IL, USA.

Peterson, K., Deeduvanu, R., Kanjamala, P., & Boles, K. (2016). *A blockchain-based approach to health information exchange networks.* U.S. Department of Health and Human Services. *https://www.healthit.gov/sites/default/files/12-55-blockchain-based-approach-final.pdf*

Puel, A., Wangenheim, A. V., Meurer, M. I., & de Macedo, D. D. J. (2014). BUCOMAX: Collaborative multimedia platform for real-time manipulation and visualization of bucomaxillofacial diagnostic images. In *2014 IEEE 27th international symposium on computer-based medical systems* (pp. 392–395). *https://doi.org/10.1109/CBMS.2014.12*

Randall, D., Goel, P., & Abujamra, R. (2017). Blockchain applications and use cases in health information technology. *Journal of Health & Medical Informatics, 8*(3), 1–17. *https://doi.org/10.4172/2157-7420.1000276*

Rayhman, M. A., Hossain, M. S., Islam, M. S., Alrajeh, N. A., & Muhammad, G. (2020). Secure and provenance enhanced internet of health things framework: A blockchain managed federated learning approach. *IEEE Access, 8*, 205071–205087. *https://doi.org/10.1109/ACCESS.2020.3037474*

Robertson, A., Cresswell, K., Takian, A., Petrakaki, D., Crowe, S., Cornford, T., Barber, N., Avery, A., Fernando, B., Jacklin, A., Prescott, R., Klecun, E., Paton, J., Lichtner, V., Quinn, C., Ali, M., Morrison, Z., Jani, Y., Waring, J., Marsden, K., & Sheikh, A. (2010). 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*, Article c4564. *https://doi.org/10.1136/bmj.c4564*

Roehrs, A., da Costa, C. A., & da Righi, R. R. (2017). Omni**PHR**: A distributed architecture model to integrate personal health records. *Journal of Biomedical Informatics, 71*, 70–81. *https://doi.org/10.1016/j.jbi.2017.05.012*

Samuel, A. M., & Garcia-Constantino, M. (2022). User-centred prototype to support wellbeing and isolation of software developers using smartwatches. *Advances in Notes in Information Science, 1*, 140–151. *https://doi.org/10.47909/anis.978-9916-9760-0-5.125*

Samuel, R. E. (2016). A layered architectural approach to understanding distributed cryptographic ledgers. *Issues in Information Systems, 17*(IV), 222–226. *https://doi.org/10.48009/4_iis_2016_222-226*

Schauz, D. (2014). What is basic research? Insights from historical semantics. *Minerva, 52*(3), 273–328. *https://doi.org/10.1007/s11024-014-9255-0*

Sembay, M. J., de Macedo, D. D. J., & Dutra, M. L. (2021). A proposed approach for provenance data gathering. *Mobile Networks and Applications, 26*(1), 304–318. *https://doi.org/10.1007/s11036-020-01648-7*

Sembay, M. J., de Macedo, D. D. J., Júnior, L. P., Braga, R. M. M., & Sarasa-Cabezuelo, A. (2023). Provenance data management in health information systems: A systematic literature review. *Journal of Personalized Medicine, 13*(6), 991. *https://doi.org/10.3390/jpm13060991*

Sembay, M. J., de Macedo, D. D. J., & Marquez Filho, A. A. G. (2022). Identification of the relationships between data provenance and blockchain as a contributing factor for health information systems. In *Proceedings of data and information in online environments: third eai international conference, DIONE 2022* (pp. 258–272). Springer Nature Switzerland. *http://doi.org/10.1007/978-3-031-22324-2_20*

Sembay, M. J., & Macedo, D. D. J. (2022). Health information systems: proposal of a provenance data management methodinthe instantiation of the W3C PROV-DM model. *Advances in Notes in Information Science, 2*, 101. ColNes Publishing. *https://doi.org/10.47909/anis.978-9916-9760-3-6.101*

Sembay, M. J., Macedo, D. D., & Dutra, M. L. (2020). 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)* (pp. 1–14). EAI. *https://doi.org/10.1007/978-3-030-50072-6_8*

Silva, P. P. da, Silva, D., McGuinness, D. L., & McCool, R. (2003). Knowledge provenance infrastructure. *IEEE Data Engineering Bulletin, 26*(4), 26–32. *https://dspace.rpi.edu/items/cd532a33-7392-4046-a4a2-c71679ec66eb*

Simmhan, Y. L., Plale, B., & Gannon, D. (2005). *A survey of data provenance techniques* [Technical Report No. TR-618]. Computer Science Department, Indiana University. *https://legacy.cs.indiana.edu/ftp/techreports/TR618.pdf*

Sligo, J., Gauld, R., Roberts, V., & Villac, L. (2017). A literature review for large-scale health information system project planning, implementation and evaluation. *International Journal of Medical Informatics, 97*, 86–97. *https://doi.org/10.1016/j.ijmedinf.2016.09.007*

Sultan, K., Ruhi, U., & Lakhani, R. (2018). *Conceptualizing blockchains: Characteristics & applications*. arXiv. *https://arxiv.org/abs/1806.03693*

Swan, M. (2015). *Blockchain: Blueprint for a new economy*. O'Reilly Media.

Tan, W. C. (2008). Provenance in databases: Past, current, and future. *IEEE Data Engineering Bulletin, 30*(4), 3–12. *https://scispace.com/pdf/provenance-in-databases-past-current-and-future-ymbe17g99v.pdf*

Tian, F. (2016). *An agri-food supply chain traceability system for China based on RFID & blockchain technology.* In *13th International conference on service systems and service management (ICSSSM)* (pp. 1–6). IEEE. *https://doi.org/10.1109/ICSSSM.2016.7538424*

Weerakoon, B. S., & Chandrasiri, N. R. (2023). Knowledge and utilisation of information and communication technology among radiographers in a lower-middle-income country. *Radiography*, *29*(1), 227–233. *https://doi.org/10.1016/j.radi.2022.11.013*

Werder, K., Ramesh, B., & Zhang, R. (2022). Establishing data provenance for responsible artificial intelligence systems. *ACM Transactions on Management Information Systems (TMIS)*, *13*(2), 1–23. *https://doi.org/10.1145/3503488*

World Health Organization (WHO). (2004). *Developing health management information systems: A practical guide for developing countries.* World Health Organization Regional Office for the Western Pacific. *https://iris.wpro.who.int/handle/10665.1/5498*.

World Health Organization (WHO). (2008). *Framework and standards for country health information systems* (2nd ed.). *https://www.who.int/healthinfo/country_monitoring_evaluation/who-hmn-framework-standards-chi.pdf*.

Zhang, J., Sun, J., & Stahl, J. N. (2003). PACS and web-based image distribution and display. *Computerized Medical Imaging and Graphics*, *27*(2–3), 197–206. *https://doi.org/10.1016/S0895-6111(02)00074-5*

Zhang, P., White, J., Schmidt, D. C., & Lenz, G. (2017). Blockchain technology use cases in healthcare. *Advances in Computers, 111*, 1–41. https:// *doi.org/10.1016/bs.adcom.2018.03.006*

Zhang, P., White, J., Schmidt, D. C., Lenz, G., & Rosenbloom, S. T. (2018). FHIRChain: Applying blockchain to securely and scalably share clinical data. *Computational and Structural Biotechnology Journal, 16*, 267–278. *https://doi.org/10.1016/j.csbj.2018.07.004*