Enhancing Data Provenance in mHealth: An Architecture Integrating FHIR with Blockchain

Gislainy Crisostomo Velasco
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
gislainycrisostomo@discente.ufg.br

Noeli Antônia Pimentel Vaz
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
Universidade Estadual de Goiás
Anapólis, Goiás, Brazil
noeli@ueg.br

Marcos Alves Vieira
Instituto Federal Goiano
Iporá, Goiás, Brazil
marcos.vieira@ifgoiano.edu.br

Sergio T. Carvalho
Universidade Federal de Goiás
Goiânia, Goiás, Brazil
sergiocarvalho@ufg.br

ABSTRACT

The advancement of digitalization in healthcare, driven by mobile health (mHealth) applications, has expanded the collection and sharing of clinical data, making information traceability a significant challenge. Data provenance mechanisms are essential to record the origin, modifications, and access to information, thereby promoting transparency, trustworthiness, and security. Although models such as W3C PROV and FHIR Provenance exist, their adoption in mHealth remains limited, primarily due to the lack of compatible architectures and the scarcity of specific technical knowledge. This paper proposes a client-server architecture compatible with W3C PROV and FHIR Provenance, integrating a Provenance Gateway responsible for transforming clinical data into FHIR resources enriched with provenance information. Hyperledger Fabric also stores traceability metadata, ensuring immutability, integrity, and auditability without exposing sensitive data. Preliminary results indicate that the proposed approach enables integrating existing mHealth applications with data provenance mechanisms, promoting interoperability and control over information traceability.

KEYWORDS

Data Provenance, W3C PROV, mHealth, Blockchain

1 INTRODUCTION

The advancement of digitalization in the healthcare sector has significantly expanded the possibilities for collecting, analyzing, and sharing clinical data. Mobile health (mHealth) applications represent one of the main drivers of this transformation, enabling continuous patient monitoring, support for therapeutic adherence, and the personalization of treatments [5, 14].

In highly dynamic and distributed contexts, such as those of mHealth applications, the traceability of clinical information presents both technical and conceptual challenges. In this scenario, data provenance mechanisms play a central role by enabling the recording of the origin, modifications, and access to information over time, thereby promoting greater transparency, reliability, and security in the use of health data [11]. The adoption of provenance supports the validation of data integrity and interoperability among heterogeneous systems.

Although standardized models such as W3C PROV and FHIR Provenance provide well-established guidelines for representing data provenance, their adoption in mHealth systems remains limited. Developers face concrete challenges, including the lack of architectures compatible with traceability metadata management and the insufficient technical expertise to correctly implement these models [2, 4].

The technical complexity in adapting formal traceability models to the architectures commonly employed in mHealth applications and the performance limitations inherent to mobile devices have hindered the widespread adoption of provenance mechanisms in this domain. These applications predominantly adopt a client-server architecture, in which mobile devices function as clients and interact with central servers through RESTful APIs [4]. This widely established pattern is valued for its simplicity, scalability, and compatibility across multiple platforms. Proposing an architecture compatible with this model – capable of incorporating provenance metadata in a transparent, structured, and interoperable manner – represents a fundamental step toward enabling traceability in mHealth applications.

The present paper proposes an architecture based on the client-server model, compatible with W3C PROV and FHIR Provenance standards. This enables the structured and interoperable integration of provenance metadata. The proposal includes an intermediate layer – the *Provenance Gateway* –that is responsible for transforming clinical data recorded by the application into FHIR resources enriched with provenance information. Additionally, the architecture incorporates integration with a blockchain network based on Hyperledger Fabric, which is dedicated to storing traceability metadata, thereby ensuring its immutability, integrity, and auditability without compromising the confidentiality of sensitive data. The proposed solution aims to facilitate the adoption of provenance mechanisms by developers, including in legacy systems, contributing to enhanced traceability and data quality within the mHealth ecosystem.

This paper is organized as follows: Section 2 presents the concepts of blockchain and data provenance, the FHIR standard, and the definition of mHealth. Section 3 reviews related studies on the application of provenance in different digital health contexts. Section 4 describes the proposed architecture. Section 5 discusses the challenges and potential of using data provenance in mHealth applications. Finally, Section 6 presents the study's main conclusions and suggests directions for future work.

2 BACKGROUND

2.1 Blockchain

Integrating blockchain and data provenance offers significant benefits from technical and organizational perspectives [19]. The immutability feature ensures that data cannot be altered without network consensus once recorded, thereby preserving its historical integrity [13]. Transparency and auditability enable stakeholders to verify data lineage reliably [12]. Decentralization eliminates single points of failure, increasing the system's resilience to faults and attacks. Trust is established through combining cryptographic mechanisms, distributed consensus, and immutability [19]. Data origin and ownership traceability is securely ensured, allowing for monitoring modifications and transfers over time.

The relevance of the model based on the integration of blockchain and data provenance stems from several factors. First, this model addresses the growing demand for guarantees of data integrity and reliability, which are essential for decision-making, regulatory compliance, audits, and accountability mechanisms [10, 13, 19]. Second, it enables the development of trustworthy data ecosystems by facilitating collaboration among entities without pre-established trust relationships. The traceability of data origins and modifications also contributes to transparency. Finally, the reduced reliance on trusted intermediaries represent a structural advantage made possible by the decentralized trust enabled by blockchain technology [10, 19].

2.2 Data Provenance

Data provenance refers to the systematic recording of the origin, transformations, and trajectory of a data item throughout its lifecycle [1, 8, 18]. It functions as contextual metadata, documenting the agents responsible for generating the information, the transformations applied, and the contexts in which the data were used [20].

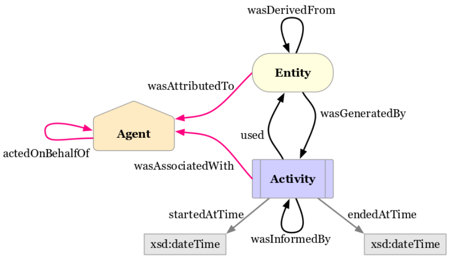


Figure 1: Conceptual model of the W3C PROV ontology.

The W3C PROV ontology (Figure 1) was developed to provide a structured and standardized model for representing data provenance and enable a detailed description of how data are generated, modified, and used over time. Its structure is based on the interaction among three main components: entity (*prov:Entity*), which represents any resource or data whose provenance can be traced; activity (*prov:Activity*), which describes the processes responsible for the creation, modification, or use of entities; and agent (*prov:Agent*), which refers to the actors accountable for these activities, including individuals, organizations, or automated systems.

2.3 FHIR

Fast Healthcare Interoperability Resources (FHIR), developed by HL7, is a standard designed to promote interoperability among health information systems through RESTful APIs and structured formats such as JSON and XML, enabling the secure and standardized exchange of clinical data across different platforms [3]. Its resources represent clinical and administrative entities, organizing essential information for healthcare delivery.

Among these resources, FHIR Provenance[1] enables the tracking of the origin and history of clinical data by recording information about the author of the action, the time it was performed, and the corresponding context. This resource adopts the 5Ws model (*Who, What, When, Where, Why*), contributing to clinical records' completeness, reliability, and transparency.

2.4 mHealth

The term mHealth refers to the practice of medicine and public health supported by mobile devices [15, 17]. This concept involves using wireless technology and mobile devices to provide healthcare services [15]. Regarded as an evolution of telemedicine, mHealth employs wireless mobile devices and communication technologies to deliver medical assistance services [7]. Mobile computing, medical sensors, and communication technologies applied to healthcare are encompassed by mHealth [9]. It is a subset of eHealth, distinguished by the use of mobile devices for delivering services to patients. While eHealth focuses on Internet-based platforms, mHealth emphasizes using wireless mobile technologies [9].

The design and development of mHealth applications face significant challenges due to the need for compatibility with mobile devices that have varying hardware specifications [4]. Services in mHealth are often developed through ad hoc approaches, employing customized software architectures and distinct processing components, which lead to incompatible designs and fragmented solutions [4].

Some mHealth applications adopt a centralized or client-server architecture [4]. Additionally, developers' lack of knowledge and experience in implementing security measures for mHealth applications constitutes a significant barrier [2]. The absence of standardization remains an obstacle [2, 4, 16], contributing to the fragmentation of solutions and incompatibility between different systems.

Given these concerns, adding provenance further intensifies the existing challenges in developing mHealth applications. Tracking the origin and usage of data would require more sophisticated architectures, increasing system complexity. Additionally, developers would need to acquire specialized knowledge of models such as W3C PROV and FHIR Provenance.

3 RELATED WORK

Data provenance has been extensively investigated in digital health, covering aspects ranging from the traceability of electronic health

[1] https://build.fhir.org/provenance.html

records (EHR) to data auditing and security in mHealth applications. Various approaches have been proposed to model and ensure such data's integrity, authenticity, and interoperability, employing ontology-based models, such as W3C PROV and FHIR Provenance, and solutions integrating blockchain and cloud computing. This section reviews the related works and compares their proposals with the approach presented in this study, which focuses on applying provenance for mHealth based on W3C PROV and FHIR Provenance standards.

Can and Yilmazer [6] propose an ontology-based provenance management model focused on enhancing privacy in healthcare systems, enabling detailed tracking of medical data history. Aiming to overcome the limitations of existing standards, the work [21] proposes a model based on FHIR Provenance and W3C PROV, seeking to expand the ability of these standards to represent the inherent complexity of health data. The authors highlight challenges related to privacy, security, regulatory compliance, and traceability, emphasizing the need for improvements to meet the specific requirements of the digital health domain.

Integrating blockchain technology with provenance mechanisms has been explored in various contexts to ensure data transparency, immutability, and auditability. In cloud computing environments, frameworks such as Blockcloud [19] and ProvChain [13] demonstrate how user-performed operations – such as data creation, modification, deletion, and sharing – can be recorded on the blockchain, forming a verifiable history of actions. In IoT devices, solutions such as BlockPro [10] use hash or metadata records stored on the blockchain to ensure data integrity from the point of collection.

Despite the identified advances, there remains a significant gap in the integrated application of provenance mechanisms in mHealth solutions that combine W3C PROV and FHIR Provenance standards with blockchain technology. This gap underpins and motivates the approach proposed in this study.

4 AN ARCHITECTURE INTEGRATING FHIR WITH BLOCKCHAIN

To ensure the traceability, integrity, and auditability of clinical information in mobile health applications, a hybrid architecture is proposed (Figure 2, on the next page), integrating conventional data repositories, interoperability mechanisms based on the FHIR standard, and a blockchain component dedicated exclusively to the storage of provenance metadata. This approach aims to meet the increasing demands for transparency and regulatory compliance in digital health environments without compromising system efficiency and scalability.

The architecture consists of distinct functional layers. The client interfaces – which may include native, hybrid, or web applications – are responsible for user interaction, data collection, and the transmission of information through RESTful APIs. The server processes business rules, authentication, and validations, and coordinates the data flow among the other components, serving as the application's operational core.

Structured data generated by users, such as activity logs, forms, or clinical notes, is stored in a traditional database, which may be relational or non-relational, depending on the application's specific requirements. This layer ensures adequate performance and backward compatibility while maintaining flexibility regarding the adopted data model.

Between the database and the interoperability repositories lies the *Provenance Gateway*, an intermediate module responsible for converting clinical data into the FHIR standard and automatically generating provenance records using the FHIR Provenance resource. This layer enables integration with other health information systems, ensuring that standardized and traceable metadata accompanies each clinical interaction.

The FHIR server stores clinical data in compliance with international health interoperability standards, facilitating the secure sharing of information with electronic health records, research platforms, and public health networks. Provenance records – containing information such as authorship, date, motivation, and context of changes – are structured using the FHIR Provenance format, ensuring compatibility with interoperable and auditable ecosystems.

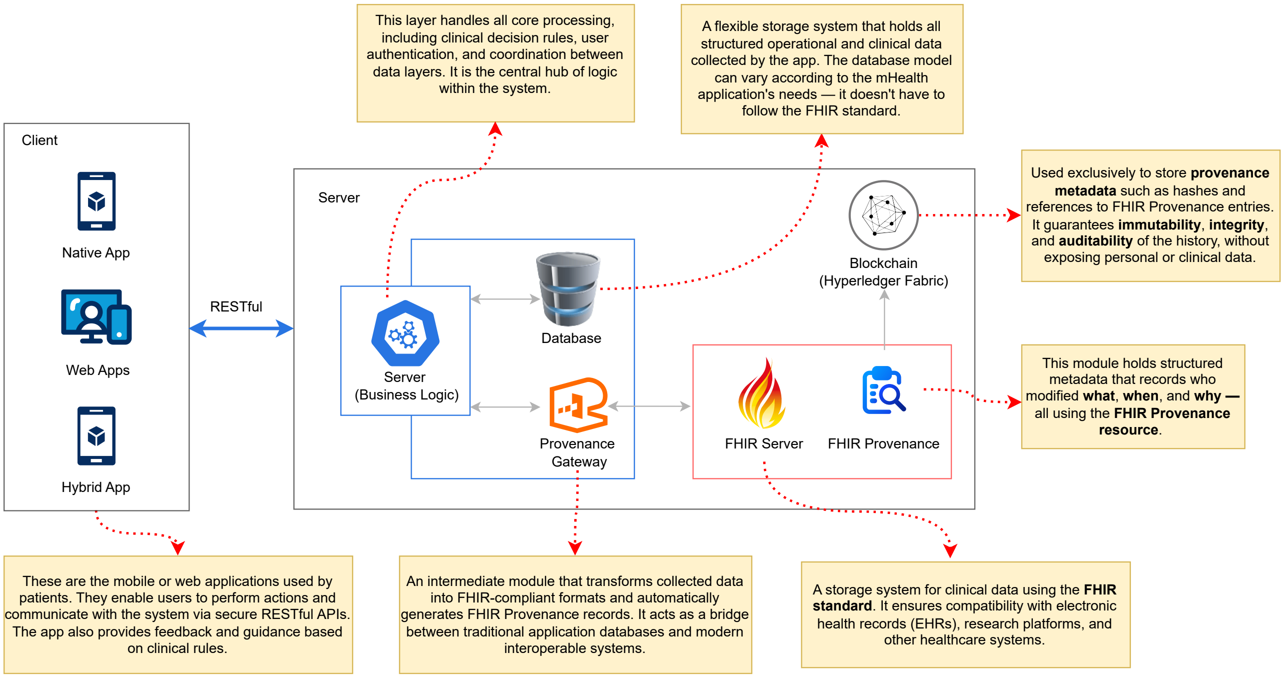
The blockchain layer, implemented with permissioned technologies such as Hyperledger Fabric, is used exclusively to record provenance metadata, including cryptographic hashes and reference identifiers to records stored on the FHIR server. This approach ensures the immutability and verifiability of recorded events without directly exposing sensitive patient data. The separation between clinical data and traceability metadata balances security and privacy requirements and system performance and scalability.

By restricting the use of blockchain to provenance data, the processing and storage overhead associated with fully decentralized solutions is avoided while incorporating an additional layer of trust. This approach is particularly relevant for mHealth applications that require integrity in clinical records, interoperability with third-party systems, and compliance with regulations such as the General Data Protection Regulation (GDPR) in the European Union, the Health Insurance Portability and Accountability Act (HIPAA) in the United States, and the General Data Protection Law (LGPD) in Brazil. The proposed architecture is generic and can be adapted to various contexts of use within the domain of mHealth applications.

5 RESULTS AND DISCUSSION

The architecture proposed in this study was designed to meet the specific requirements of mHealth applications, which operate in environments characterized by technical limitations, device heterogeneity, and high demand for continuity in patient-centered care. These applications play a significant role in health promotion, chronic condition monitoring, and treatment personalization; however, they face recurring challenges related to interoperability, traceability of clinical information, and compliance with data protection regulations. In this context, the proposed architecture represents a technically viable alternative to enhance reliability and foster the integration of these solutions within the broader digital health ecosystem.

One of the main distinguishing features of the proposed approach lies in its ability to integrate with existing systems without requiring significant modifications to the data infrastructure. Maintaining operational databases – relational or non-relational – as the foundation for clinical and functional storage allows mHealth applications to continue operating without interruption while simultaneously enabling the generation of interoperable and auditable records. This



This layer handles all core processing, including clinical decision rules, user authentication, and coordination between data layers. It is the central hub of logic within the system.

A flexible storage system that holds all structured operational and clinical data collected by the app. The database model can vary according to the mHealth application's needs — it doesn't have to follow the FHIR standard.

Used exclusively to store **provenance metadata** such as hashes and references to FHIR Provenance entries. It guarantees **immutability**, **integrity**, and **auditability** of the history, without exposing personal or clinical data.

This module holds structured metadata that records who modified **what**, **when**, and **why** — all using the **FHIR Provenance resource**.

These are the mobile or web applications used by patients. They enable users to perform actions and communicate with the system via secure RESTful APIs. The app also provides feedback and guidance based on clinical rules.

An intermediate module that transforms collected data into FHIR-compliant formats and automatically generates FHIR Provenance records. It acts as a bridge between traditional application databases and modern interoperable systems.

A storage system for clinical data using the **FHIR standard**. It ensures compatibility with electronic health records (EHRs), research platforms, and other healthcare systems.

Figure 2: Proposed architecture for mHealth applications integrating FHIR-based interoperability and blockchain-based provenance. It preserves existing data structures while enabling secure and auditable clinical data tracking.

approach supports incremental adoption, respecting each application's technical and operational particularities, and significantly reduces the costs and efforts associated with migration to advanced interoperability standards.

The inclusion of the Provenance Gateway in the architecture has proven, in initial tests, to be an effective strategy for enabling the automated generation of traceability metadata based on the FHIR Provenance standard. This functionality allows applications originally developed without interoperability support to begin producing structured data ready to integrate with electronic health records and other clinical platforms. The standardization of provenance records enhances auditability and contributes to the security and transparency of using clinical information.

Another key component is adopting blockchain technology, specifically through the permissioned platform Hyperledger Fabric. This platform is employed exclusively for storing provenance metadata, such as cryptographic hashes and record identifiers. Hyperledger Fabric adds critical functionalities for clinical environments, including access control, modularity, and support for private transactions among participating institutions.

Based on ongoing implementations, preliminary results indicate that the proposed architecture offers an appropriate balance between flexibility, interoperability, and security, positioning it as a viable alternative for modernizing legacy mHealth applications and developing new solutions aligned with contemporary technical and regulatory standards. By enabling clinical traceability with integrity assurance without compromising scalability and privacy, the proposal significantly contributes to strengthening data governance in digital health.

6 CONCLUSIONS

The proposed architecture represents a significant advancement in adopting data provenance mechanisms in mHealth applications from both technical and practical perspectives. Its modular structure enables overcoming persistent challenges in the sector, such as the lack of standardization, the complexity of integration with large-scale systems, and the growing demand for auditability and regulatory compliance. By enabling the evolution of existing solutions and providing a robust technical foundation for developing new applications, the architecture strengthens clinical data governance, fosters greater transparency in data-driven decisions, and promotes consolidating a more resilient and integrated digital health ecosystem.

As a continuation of this research, upcoming efforts are directed toward evaluating the architecture in different mHealth application scenarios, focusing on existing solutions. The aim is to analyze its performance, scalability, and compliance with these applications' actual technical and functional requirements to validate its applicability in diverse operational contexts.

Another relevant direction for future work involves exploring the practical use of data provenance in the specific context of mHealth applications. Beyond the structured storage of information, it is essential to understand how provenance records can be integrated into the workflows of healthcare professionals using mobile devices.

Future investigations should consider the potential of these data to support real-time clinical decision-making, personalize interventions, and enhance the patient experience in mobile environments.

REFERENCES

[1] Mansoor Ahmed, Amil Rohani Dar, Markus Helfert, Abid Khan, and Jungsuk Kim. 2023. Data Provenance in Healthcare: Approaches, Challenges, and Future Directions. *Sensors* 23, 14 (2023). https://doi.org/10.3390/s23146495

[2] Bakheet Aljedaani et al. 2021. Challenges with developing secure mobile health applications: systematic review. *JMIR mHealth and uHealth* 9, 6 (2021), e15654.

[3] Muhammad Ayaz et al. 2021. The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities. *JMIR medical informatics* 9, 7 (2021), e21929.

[4] Ana González Bermúdez et al. 2024. A fusion architecture to deliver multipurpose mobile health services. *Computers in Biology and Medicine* 173 (2024), 108344.

[5] Blind Peer Review. 2024. Blind peer review. (2024).

[6] Ozgu Can and Dilek Yilmazer. 2020. Improving privacy in health care with an ontology-based provenance management system. *Expert Systems* 37, 1 (2020), e12427.

[7] Faizel Faker. 2018. *Mobile Health Data: Investigating the data used by an mHealth app using different mobile app architectures.* Ph. D. Dissertation. Department of Computer Science. http://hdl.handle.net/11427/31242

[8] Yolanda Gil et al. 2010. Provenance XG final report. (2010).

[9] Robert SH Istepanian. 2022. Mobile health (m-Health) in retrospect: the known unknowns. *International journal of environmental research and public health* 19, 7 (2022), 3747.

[10] Uzair Javaid, Muhammad Naveed Aman, and Biplab Sikdar. 2018. Blockpro: Blockchain based data provenance and integrity for secure iot environments. In *Proceedings of the 1st workshop on blockchain-enabled networked sensor systems*. 13–18.

[11] Marco Johns, Lena Baum, and Fabian Prasser. 2025. Tracking provenance in clinical data warehouses for quality management. *International Journal of Medical Informatics* (2025). https://doi.org/10.1016/j.ijmedinf.2024.105690

[12] Tsung-Ting Kuo, Hyeon-Eui Kim, and Lucila Ohno-Machado. 2017. Blockchain distributed ledger technologies for biomedical and health care applications. *Journal of the American Medical Informatics Association* 24, 6 (2017), 1211–1220.

[13] Xueping Liang, Sachin Shetty, Deepak Tosh, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla. 2017. Provchain: A blockchain-based data provenance architecture in cloud environment with enhanced privacy and availability. In *2017 17th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGRID)*. IEEE, 468–477.

[14] Richard K. Lomotey and Ralph Deters. 2014. Mobile-Based Medical Data Accessibility in mHealth. In *2014 2nd MobileCloud*. 91–100. https://doi.org/10.1109/MobileCloud.2014.24

[15] Dounia Marbouh et al. 2022. A blockchain-based regulatory framework for mHealth. *Data* 7, 12 (2022), 177.

[16] Mario Nacinovich. 2011. Defining mHealth. , 3 pages.

[17] Simon P Rowland et al. 2020. What is the clinical value of mHealth for patients? *NPJ digital medicine* 3, 1 (2020), 4.

[18] Marcio Jose Sembay. 2023. *PROV-Health: método para gerenciamento de dados de proveniência em sistemas de informação em saúde.* Ph. D. Dissertation. UFSC. https://repositorio.ufsc.br/handle/123456789/251582

[19] Sachin Shetty, Val Red, Charles Kamhoua, Kevin Kwiat, and Laurent Njilla. 2017. Data provenance assurance in the cloud using blockchain. In *Disruptive Technologies in Sensors and Sensor Systems*, Vol. 10206. SPIE, 125–135.

[20] Marten Sigwart et al. 2020. A secure and extensible blockchain-based data provenance framework for the Internet of Things. *Personal and Ubiquitous Computing* (2020), 1–15.

[21] Maria Judith Velez De Villa Rojas. 2024. *Provenance information in eHealth.* Ph. D. Dissertation. UPC, Facultat d'Informàtica de Barcelona, Departament d'Arquitectura de Computadors. http://hdl.handle.net/2117/419443