The Role of Blockchain Technology in Ensuring Security and Immutability of Open Data in Healthcare

Tiago Guimarães[1[0000-0001-6320-8878]], Ricardo Duarte[1[0000-0001-6423-3442]] João Cunha[1[0000-0001-7120-5706]], Pedro Silva[1[0009-0003-7926-1587]] and Manuel Filipe Santos[1[0000-0002-5441-3316]]

[1]Algoritmi Reasarch Center, School of Engineering, University of Minho, Azurém Campus, Guimarães, 4800-05, Portugal
tsg@dsi.uminho.pt

Abstract. Clinical information is highly confidential due to its sensitive nature. Implementing health information systems has raised concerns regarding interoperability, privacy, and security. The storage and retrieval of this information also present the same problems. Therefore, any effort to introduce healthcare information systems must ensure patient data's safety, privacy, integrity, and immutability. Blockchain technology and the openEHR open data model have emerged to address these concerns, providing a solution that guarantees data security, interoperability between systems, and the accuracy of stored data queries. Two different architectures were developed and subjected to several performance tests to enhance security and immutability in open data models implemented in healthcare institutions. The results were analysed to determine which architecture provides more value to a healthcare institution. Subsequently, a discussion was held to draw appropriate conclusions.

Keywords: Blockchain, Blockchain in Healthcare, OpenEHR, Benchmarking.

1 Introduction

The health sector is a sector that has unique requirements, and trusting the data from its activities is essential for its operations [1], [2]. With the growing amount of data generated, some problems arise, including unauthorised sharing, invasions, and theft of confidential data. These propensities lead to people's suspicions and doubts about the trust veracity of these institutions [2], [3]. It is essential to consider alternative approaches, like blockchain technology, to address these issues. Given its nature and characteristics, it offers a solution to the needs demanded by the sector [4], [5]. Allied with this technology comes the open data structure, openEHR. It enables reliable structuring, management, storage, and patient data integration across healthcare organisations. The main idea is to standardise concepts related to health used in databases or Electronic Health Record (EHR) systems in a set of libraries called archetypes [6].

The present work is divided into several sections, starting with a brief introduction and then a literature review about OpenEHR, blockchain and blockchain in healthcare. Section five is discussed why to use openEHR with blockchain. The fol-

lowing section presents the developed architectures, which will serve as the basis for the tests performed in section seven. Finally, section eight is presented a discussion of the results obtained and the conclusions, as well as future work tracing.

2 What is OpenEHR?

OpenEHR is an open-source, vendor-neutral standard for storing, retrieving, and exchanging electronic health records (EHRs). It was first developed in 2000 by a consortium of international health informatics experts to create a shared model for EHR systems that could be used across different healthcare settings and countries. The OpenEHR approach is based on archetypes and flexible and reusable templates for capturing clinical information. These archetypes can be combined and modified to create specific EHR instances that meet the needs of different healthcare organisations and individual patients [1].

OpenEHR is designed to overcome the limitations of traditional EHR systems, which are often siloed, proprietary, and difficult to customise. Using a standardised model for clinical data, OpenEHR enables interoperability between different EHR systems and between EHRs and other health information systems, such as decision support tools and population health management platforms. OpenEHR also supports open APIs and web services, allowing seamless integration with other healthcare applications and data sources [2]–[4].

3 Blockchain

Blockchain technology is a distributed and secure database that enables the recording of transactions in a reliable and immutable manner. Each transaction is validated by a network of peers, making data manipulation or falsification harder [11], [12].
According to Swan (2015) [11], blockchain can be defined as a "distributed digital ledger that uses cryptography to maintain secure and verifiable records." The ledger is composed of blocks, which contain information about recent transactions. Each block is connected to the previous block and is verified by a distributed network of computers.

According to Tapscott and Tapscott (2016) [12], blockchain is "a new technological platform that can help improve the way we record and manage data." Through blockchain, data exchange can be done without intermediaries or trusted third parties, allowing for a more secure and efficient transfer of value.
Blockchain is primarily known as the technology behind cryptocurrencies, but its potential for use extends far beyond that, including data management, supply chains, and electronic voting [11], [12].

4 Blockchain in the Healthcare Sector

Blockchain is an emerging technology that has gained prominence in various fields, including healthcare. It is a promising technology that can transform healthcare by providing a secure and efficient platform to store and share confidential medical information [13], [14]. According to de Hasselgren et al. (2019) [2], blockchain can provide an immutable and transparent record of health information, making sharing medical information more straightforward and secure.

Another significant benefit of using blockchain technology in the healthcare industry is data security. The blockchain can store medical data in an encrypted and decentralised manner, reducing the risk of data breaches and ensuring patient privacy. Furthermore, blockchain can help improve the efficiency of exchanging medical information between different healthcare providers and reduce medical errors [15] [16].
According to some authors, blockchain can help improve the transparency and efficiency of clinical trials and ensure the authenticity of collected data. Furthermore, blockchain can create immutable records of genomes and health data, which can be used to develop personalised treatments based on each patient's genetic characteristics [17], [18]. In addition to the above, several ongoing initiatives to use blockchain in healthcare include blockchain-based electronic medical record systems, health data sharing platforms, blockchain-based patient identity management systems, and drug tracking [19], [20].

Blockchain technology can potentially transform the healthcare industry, but its adoption still faces many challenges. It is necessary to work with stakeholders to overcome these obstacles and ensure that blockchain technology is safe, efficient, and properly regulated for use in healthcare [13], [21].

In this case of the study, to guarantee the guarantee the privacy and safety of patient’s records, we are going utilize the permissioned blockchains. A permissioned blockchain is a system where the identity of the entities is controlled by an identity provider. The latter is responsible to maintain and control network access and user’s participation in consensus and block validation [5].

5 Why openEHR with blockchain in healthcare?

OpenEHR is a standard for electronic health record (EHR) systems that allows for creation of interoperable and vendor-neutral health data. Blockchain technology can enhance the security, privacy, and interoperability of OpenEHR by providing a decentralised and tamper-resistant platform for storing and sharing health data. By combining OpenEHR with blockchain, healthcare providers can create a secure and transparent environment for managing patient health data, improving patient outcomes, and reducing costs. Furthermore, blockchain technology can also support the creation of a patient-centric healthcare system, where patients have control over their health data and can choose who can access it [1], [6], [7].

In order to understand the Strengths, Weaknesses, Opportunities, and Threats (SWOT) in the combined use of openEHR with blockchain in healthcare, a SWOT analysis was conducted based on the publication of various authors.

The **strengths** highlighted by the authors are that OpenEHR provides a structured data model based on standards that can improve interoperability and data integration in healthcare [8]. Blockchain can enhance the security and privacy of healthcare data, ensuring that only authorised individuals can access sensitive information. The combination of OpenEHR with blockchain can provide an open, decentralised, and trustworthy solution for managing health data [6], and the combination of OpenEHR with blockchain technology can enhance the security of electronic health records, protecting patient data from breaches and unauthorised access [3].

The **weakness**es highlighted by the authors are that implementing OpenEHR and blockchain in healthcare may be expensive and require a high investment in information technology (IT) infrastructure and resources [9]. Implementing OpenEHR requires a certain level of technical expertise and may be complex for end users [1]. The use of blockchain has yet to be widely adopted in the healthcare industry, still facing significant challenges for its implementation, which may limit its effectiveness in certain areas [10].

The **opportunities** highlighted by the authors include the integration of OpenEHR with blockchain can provide a technical foundation for decentralised healthcare applications, such as digital health wallets and health tracking apps [11], [12]. Blockchain-based solutions can increase public trust in digital health systems and promote wider adoption of digital health technologies [13]. The growing demand for health data security and privacy protection may drive the adoption of more advanced technologies, such as OpenEHR with blockchain [14].

The **threats** highlighted by the authors include that data privacy may be compromised if blockchain technology is not implemented correctly or privacy policies are insufficient [10]. Resistance to change by healthcare professionals and patients may limit the adoption of digital health technologies, including solutions based on blockchain and openEHR [15], and the adoption of emerging technologies such as OpenEHR with blockchain may face regulatory and compliance challenges [10].

Combining OpenEHR with blockchain technology can offer numerous benefits to the healthcare industry, including the security of electronic medical records, the protection of patient data against violations and unauthorised access, patient control over their health data, and improvements in clinical research and population health. While adopting this technology may require a sophisticated IT infrastructure and changes in organisational culture, the growing demand for health data security and privacy protection can drive its implementation in the healthcare industry [3], [16].

6 Architecture

By integrating Blockchain technology with the open data model, openEHR, it is possible to ensure data interoperability and standardisation of all EHRs while maintaining

data integrity, privacy, and security. In our case study, two scenarios were considered: Scenario 1, as represented in Figure 1, and Scenario 2, as represented in Figure 2.

Analysing the flow of Scenario 1, as depicted in the following figure, begins with the insertion of clinical information and data about patients. These inputs are then transformed into EHRs and sequentially processed and analysed based on the specifications modelled in the open data model, openEHR, at the APIS layer through the gateway. The Hyperledger Caliper was used to perform a series of insertions about the patient's clinical data on the blockchain. Subsequently, the blockchain stores all the transactions in a secure, immutable, and private manner for future consultation. Finally, Hyperledger Caliper evaluates the transactions in terms of success, speed, the maximum, minimum, and average time to send and receive a response, and the average number of transactions processed per second. These metrics are presented on an HTML page.

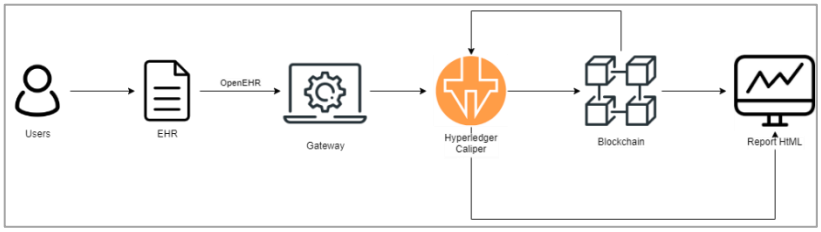

Figure 1 - First Scenario Architecture

Initiating the analysis of Scenario 2, as illustrated in the figure below, the flow begins similarly to the scenario described above. It commences with inserting clinical information for patients into the systems used by the institution. These inputs are then transformed into EHRs and sequentially processed and analysed based on the specifications modelled in the open data model, openEHR, at the APIS layer through the gateway. Continuing with the flow, a fork is observed. In this step the data comes from the same place but it will be handled and stored in different ways. Following the upper arrow in the flow, the data is stored in the hospital's database, where it can be further processed and transformed as per the needs of the hospital's stakeholders.

On the other hand, the lower path involves creating a hash block for each object through the encoding process. An object ID is associated with the block to identify the person to whom the data pertains. The main objective of this technique is to validate if there has been any intrusion or alteration to the data, thus providing enhanced security. If any changes are made to the data in the object, the MD5 will be updated, and the updated block will be stored on the blockchain. Proceeding with the flow, Hyperledger Caliper triggers the entries of the hashed objects into the blockchain. However, in this case, the constitution of the object changes, and it becomes just an ID and the hash. Next, the blockchain stores the transactions and plays a crucial role in verifying whether the recorded hash matches the updated hash. Finally, Hyperledger Caliper utilises metrics to measure performance, which are presented on an HTML page.

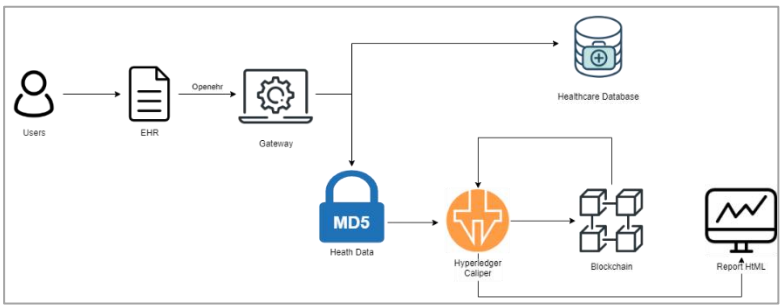

Figure 2 - Second Scenario Architecture

7 Results

In this chapter, a visual demonstration of the network's performance obtained in the various stress tests will be carried out. Thus, it will be possible to measure and understand whether the objective of increasing the security and immutability of implementing open data models in a hospital environment has been achieved. This income statement performed two tests, one for the first and the other for the second. At each test, two types of graphs are shown. One evaluates the performance of the network in the gradual submission of people. The other considers the amount of memory that the network uses in total.

7.1 Insertion of data into the blockchain with 20000 records, Scenario 1 and Scenario 2

Four insertions of 20000 people were performed on the blockchain for this test. Initially, the container was restarted. As the insertions were performed, the processing capacity of the network decreased, gradually increasing the average latency. This phenomenon is explained by the increasing amount of data stored in the blockchain. It is noteworthy that the processing speed was 5 TPS. Adding this to the physical capacity of the machine used for testing, the processing time increased considerably. A positive aspect that goes against speed stability corresponds to the absence of failures.

In this test, it is possible to see a growing increase in processing time and memory usage in both scenarios. As mentioned, the amount of data entered is twenty thousand in each of the four iterations, thus pushing the blockchain's capacity to the limit. As the iterations were carried out, the network performance naturally decreased, which caused an increase in processing time. The same occurred with memory usage, where it is possible to conclude that as the volume of data increases, the memory used by the system also increases, which will cause a slower system.

Upon analysing the two scenarios, as evidenced in Table 1, it is possible to deduce that Scenario 2 stands out with shorter processing times and memory usage in each iteration performed. Moreover, this scenario adds a layer of data security and better aligns with the internal requirements of the hospital. Considering the minimal impact on the existing implementation in the hospital setting, Scenario 2 is identified as the optimal choice. It is less resource-intensive, provides the desired benefits, and avoids disruptions. Further tests were conducted for this scenario, which will be presented in the subsequent chapters.

Table 1. Analyse between two Scenarios

		1st Interaction	2nd Interaction	3rd Interaction	4th Interaction
Process Time	**Test – Scenario 1**	583,74	688,3	768,45	1035,24
	Test – Scenario 2	60,33	312,37	442,89	763,4
Memory Usage	**Test – Scenario 1**	1679,55	2025,88	2231,38	2566,96
	Test – Scenario 2	529,39	752,26	1289,7	1631,93

7.2 Additional Test for Scenario 2

For the second test of scenario 2, a different configuration was performed, where the network defined the send rate automatically. Initially, the containers were cleared and then 1000, 5000, 10000, and 20000 people were inserted into the network. With the change made, a considerable variation in latency was observed compared to the previous tests. The achieved latency times show an improved, constant, and consistent speed. This occurs because the change allowed the system to choose the rate it could support.

Additionally, it is possible to observe that the sending speed decreases with each insertion. This evolution is expected as the network's capacity decreases as more data is inserted. It is possible to observe that memory increases progressively. There is more significant variation in memory because each iteration involves a different number of insertions, which justifies that memory is inconsistent and growing.

For the third test of scenario 2, the same configuration as test 2 was performed, where the network defined the send rate automatically. Initially, the containers were cleared, and then four insertions of 20000 people were performed each time in the network. With the changes made, considerable alterations were observed regarding latency times. The latency times increased and presented an improved, constant, and consistent speed because the change allowed the system to choose the rate it could support. Additionally, it is possible to observe that the sending speed decreased as the 20000 people were inserted. This happens because the network's capacity decreases with the continuous insertion of data, causing an overload of the network.

Regarding memory, it is possible to observe it increase progressively. In this case, memory increased consistently because the number of insertions was constant, meaning that the amount of processing by peers and orderers was practically the same, and the only thing that increased was the size of the databases. This increase is considered regular, constant, and growing because the number of records also increases. The results for both test it is visible in the following table.

Table 2 - Additional tests for Scenario 2

		1st Interaction	2nd Interaction	3rd Interaction	4th Interaction
Insertions	**Scenario 2 – Test 2**	1000	5000	10000	20000
	Scenario 2 – Test 3	20000	20000	20000	20000
Process Time	**Scenario 2 – Test 2**	3,87	3,9	4,7	4,79
	Scenario 2 – Test 3	5,16	5,4	5,14	6,38
TPS	**Scenario 2 – Test 2**	17,6	17,5	14,6	14,5
	Scenario 2 – Test 3	13,4	12,9	13,6	11,3
Memory Usage	**Scenario 2 – Test 2**	689,25	997	1139,45	1314,39
	Scenario 2 – Test 3	1566,44	1909,07	2267,47	2323,48

8 Discussion and Conclusions

Every solution requires a thorough and truthful evaluation, and engaging in an honest discussion about its efficacy is vital. In order to do so, it is helpful to utilise technology assessment methodologies like a SWOT analysis.

A SWOT analysis is a tool commonly used by organisations for managing their strategies and objectives. This tool assesses internal and external factors affecting the organisation, with SW factors being internal and OT factors being external attributes. By performing this analysis, one can determine the feasibility of implementing a given solution. The SWOT analysis yields qualitative and quantitative metrics, which can serve as indicators of the technology acceptance model as perceived by practitioners.

Strengths:

- Structured clinical data.
- Interoperability between all systems.
- Availability of data for better development of Business Intelligence and decision support systems.
- Reduction of obsolete data and poorly inserted records without clinical value and clean information analysis.

Weakness:

- Scalability and storage and processing capacity.
- Application with a real-time data update.
- Dynamics of OpenEHR structures and their versioning.
- Efficiency regarding processing time and resources.
- Relationship between OpenEHR structures and Multidimensional structures that support analytical processes considering analysis axes that cut across several patients or certain variables/characteristics of the same patient.

Opportunities:

- Standardisation of clinical records on a large scale.
- Training and integration of clinical modelling of these structures in training healthcare professionals and/or higher healthcare courses.
- Centralised data for analysis.

Threats:

- User resistance to adopting a new system by healthcare professionals.
- Negative aspects of the environment with the potential to compromise the proposed solution.
- External development of more efficient solutions.
- The emergence of a new standard with better conditions.

Integrating these technologies and methodologies in the health area is important for evolution in terms of speed of patient care and quality of service, without forgetting the importance of interoperability that OpenEHR provides for better communication between services and, consequently, BI integration.

OpenEHR models provide a disruptive way of storing data within healthcare and are primarily focused on standardising and managing clinical data, including capturing and storing structured and unstructured health information.

It is not explicitly designed for data analysis or business intelligence purposes, so current BI solutions must adapt and adjust how data is consulted efficiently.

According to our research, this integration between OpenEHR and BI has yet to be adequately investigated, so a gap in this area can be used for research. Future research work will be the realisation of artefacts to adapt a BI system in a generic way to the OpenEHR model and study the creation of a layer of Extraction Transformation and Load (ETL) that efficiently analyse data from the models.

Acknowledgements

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020

References

[1] "What is openEHR?" https://www.openehr.org/about/what_is_openehr (accessed Mar. 24, 2023).

[2] T. Beale and S. Heard, "Architecture Overview, openEHR." Foundation, 2007.

[3] J. Cunha, R. Duarte, T. Guimarães, and M. F. Santos, "Permissioned Blockchain Approach using Open Data in Healthcare," *Procedia Comput. Sci.*, vol. 210, pp. 242–247, 2022.

[4] A. Wulff, B. Haarbrandt, E. Tute, M. Marschollek, P. Beerbaum, and T. Jack, "An interoperable clinical decision-support system for early detection of SIRS in pediatric intensive care using openEHR," *Artif. Intell. Med.*, vol. 89, pp. 10–23, 2018.

[5] J. Cunha, R. Duarte, T. Guimarães, and M. F. Santos, "Permissioned Blockchain Approach using Open Data in Healthcare," *Procedia Comput. Sci.*, vol. 210, pp. 242–247, 2022, doi: 10.1016/j.procs.2022.10.144.

[6] A. Azaria, A. Ekblaw, T. Vieira, and A. Lippman, "Medrec: Using blockchain for medical data access and permission management," in *2016 2nd international conference on open and big data (OBD)*, 2016, pp. 25–30.

[7] T.-T. Kuo, H.-E. Kim, and L. Ohno-Machado, "Blockchain distributed ledger technologies for biomedical and health care applications," *J. Am. Med. Informatics Assoc.*, vol. 24, no. 6, pp. 1211–1220, 2017.

[8] D. Kalra, T. Beale, and S. Heard, "The openEHR foundation," *Stud. Health Technol. Inform.*, vol. 115, pp. 153–173, 2005.

[9] M. Benchoufi, R. Porcher, and P. Ravaud, "Blockchain protocols in clinical trials: Transparency and traceability of consent," *F1000Research*, vol. 6, p. 66, 2018.

[10] L. Soltanisehat, R. Alizadeh, H. Hao, and K.-K. R. Choo, "Technical, temporal, and spatial research challenges and opportunities in blockchain-based healthcare: A systematic literature review," *IEEE Trans. Eng. Manag.*, 2020.

[11] R. Ben Fekih and M. Lahami, "Application of blockchain technology in healthcare: A

comprehensive study," in *The Impact of Digital Technologies on Public Health in Developed and Developing Countries: 18th International Conference, ICOST 2020, Hammamet, Tunisia, June 24–26, 2020, Proceedings 18*, 2020, pp. 268–276.

[12] S. Khezr, M. Moniruzzaman, A. Yassine, and R. Benlamri, "Blockchain technology in healthcare: A comprehensive review and directions for future research," *Appl. Sci.*, vol. 9, no. 9, p. 1736, 2019.

[13] A. Hasselgren, K. Kralevska, D. Gligoroski, S. A. Pedersen, and A. Faxvaag, "Blockchain in healthcare and health sciences—A scoping review," *Int. J. Med. Inform.*, vol. 134, no. May 2019, p. 104040, 2020, doi: 10.1016/j.ijmedinf.2019.104040.

[14] S. Angraal, H. M. Krumholz, and W. L. Schulz, "Blockchain technology: Applications in health care," *Circ. Cardiovasc. Qual. Outcomes*, vol. 10, no. 9, pp. 1–4, 2017, doi: 10.1161/CIRCOUTCOMES.117.003800.

[15] T. Kumar, V. Ramani, I. Ahmad, A. Braeken, E. Harjula, and M. Ylianttila, "Blockchain utilization in healthcare: Key requirements and challenges," *2018 IEEE 20th Int. Conf. e-Health Networking, Appl. Serv. Heal. 2018*, pp. 1–7, 2018, doi: 10.1109/HealthCom.2018.8531136.

[16] G. G. Dagher, J. Mohler, M. Milojkovic, and P. B. Marella, "Ancile: Privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology," *Sustain. cities Soc.*, vol. 39, pp. 283–297, 2018.