SENSIBLE: Sequestered and Synergistic Blockchain Ecosystem

Meghana Kshirsagar[1], Gauri Vaidya Vaidya[1], Yao Yao[1], Smita Kasar[2], and Ryan Conor[1]

[1]University of Limerick
[2]Marathwada Institute of Technology

October 17, 2022

Abstract

Healthcare interoperability unfolds the way for personalized healthcare services at a reduced cost. Furthermore, a decentralized system holds the promise to prevent compromises such as cyber-attacks due to data breaches. Hence, there is a need for a framework that seamlessly integrates and shares data across the system stakeholders. We propose SENSIBLE, a blockchain-powered, knowledge-driven data-sharing framework that gives patients complete control of their medical history and can extract rich information hidden in it using knowledge graphs (KGs). By incorporating both blockchain and KGs, we can provide a platform for secure data sharing amongst stakeholders by maintaining data privacy and integrity through data authentication and robust data integration. We present a Proof-of-Concept of the SENSIBLE network with Ethereum to share dynamic knowledge across stakeholders. Dynamic knowledge generation on the blockchain provides a two-fold advantage of cooperation and communication amongst the stakeholders in the healthcare ecosystem. This leads to operational ease through sharing relevant portions of complex information while also ensuring the isolation of sensitive medical data.

RESEARCH ARTICLE

Engineering Reports

SENSIBLE: Sequestered and Synergistic Blockchain Ecosystem

Meghana Kshirsagar[1,2] | Gauri Vaidya[1,2] | Yao Yao[1,2] | Smita Kasar[3] | Conor Ryan[1,2]

[1] Biocomputing and Developmental Systems Group, University of Limerick, Limerick, Ireland

[2] Lero, The Science Foundation Ireland Research Centre for Software, Limerick, Ireland

[3] Department of Computer Science, Marathwada Institute of Technology, Maharashtra, India

Correspondence
Meghana Kshirsagar, Biocomputing and Developmental Systems Group, University of Limerick, Limerick, Ireland.
Email: meghana.kshirsagar@ul.ie

Funding information
Science Foundation Ireland, Grant/Award Number: #16/IA/4605; Science Foundation Ireland Centre for Research Training in Artificial Intelligence, Grant/Award Number: 18/CRT/6223

Healthcare interoperability unfolds the way for personalized healthcare services at a reduced cost. Furthermore, a decentralized system holds the promise to prevent compromises such as cyber-attacks due to data breaches. Hence, there is a need for a framework that seamlessly integrates and shares data across the system stakeholders. We propose *SENSIBLE*, a blockchain-powered, knowledge-driven data-sharing framework that gives patients complete control of their medical history and can extract rich information hidden in it using knowledge graphs (KGs). By incorporating both blockchain and KGs, we can provide a platform for secure data sharing amongst stakeholders by maintaining data privacy and integrity through data authentication and robust data integration. We present a Proof-of-Concept of the *SENSIBLE* network with Ethereum to share dynamic knowledge across stakeholders. Dynamic knowledge generation on the blockchain provides a two-fold advantage of cooperation and communication amongst the stakeholders in the healthcare ecosystem. This leads to operational ease through sharing relevant portions of complex information while also ensuring the isolation of sensitive medical data.

KEYWORDS
Blockchain, Ethereum Technology, Data Sharing, Data Integration, *Interoperability*, *Knowledge Graph*

1 | INTRODUCTION

Electronic Health Records (EHRs) [1] were introduced with a broader aim of improving the planning and management of health services [2]. With advancements in healthcare and a multitude of health data sources, the use of EHRs is extended to access the relevant information of the patient's medical history when required, with an objective to deliver personalized healthcare [3, 4, 5, 6]. Personalized healthcare is the detailed study of an individuals and successive treatment according to their lifestyle, health traits, and metabolism. The data in EHRs ranges from x-rays, laboratory reports, and radiology reports to allergies, immunizations, surgeries, medications, family medical history, medical bills, etc. Along with EHRs, the Internet of Things (IoT) and ubiquitous computing have led to an exponential increase in the seamless integration of internet-connected devices into our day-to-day lives generating large volumes of health-related information. According to the European Commission, around 19% of the people in the age group of 19-74 use smartwatches and fitness wearables, 11% of the people in the same age group use smart apps that track fitness activities in their daily lives, while 10% of the people use smart devices connected with IoT at home [7]. This ubiquitous integration enables the communication of these devices to track our behaviors, emotions, health, and businesses, covering even minor daily activities to draw hidden insights from them. This gives rise to the concept of multimodal data existing in the form of text, images, audio, video, etc leading to the challenge of assimilation and integration without loss of information [8, 9]. EHRs, combined with these informative data captured from multiple modalities have the potential to augment existing diagnostic systems [10]. The recent evidence also shows the low sides of incomplete EHRs, and integrated data as a major barrier in providing personalized healthcare [11]. With COVID-9, this need has been significantly increased to have a complete idea of the patient engaging in personalized and intelligent healthcare services [11] which can be attained through artificial intelligence [12]. These ideas of integration and sharing of personal data also bring in two crucial concerns: *privacy* and *security* [13].

1.1 | Motivation

There are many instances where personal information has been compromised and sometimes leaked on public sites in recent years [14]. According to the 2018 Data Breach Investigation Report [15], the contribution of healthcare data breaches is the highest among all types of breaches. Figure 1 illustrates the number of individuals affected by data breaches from the year 2010 to 2019 [16]. It can be seen from the graph that the impact was highest in the year 2015 when 113.27 million individuals were affected due to the loss of personal health records. Figure 1 also illustrates the average loss per health breach from the year 2010 to 2019. For calculating financial loss that occurred post these breaches, several factors are taken into consideration; this includes both direct and indirect expenses incurred by an organization holding the records. Thus, data breaches have an adverse effect on the privacy of individuals as well as the finance of the healthcare industry. Hence, there is a need to design a data-sharing framework for the integration of multimodal health-related data while maintaining individual ownership of data to ensure its privacy and security.

1.2 | Objectives of SENSIBLE

In this research work, we use two technologies to achieve the abovementioned objective: Blockchain [17] and Knowledge Graphs (KGs) [18]. Blockchain is a distributed public ledger that tracks each activity within the system while encrypting personal and private information with cryptographic algorithms and is widely used today to develop a trustless framework. A major advantage of blockchain is that it stores each activity on the network in the form of a transaction [17], thus making it more difficult for malicious users to tamper with the records or for hackers to steal

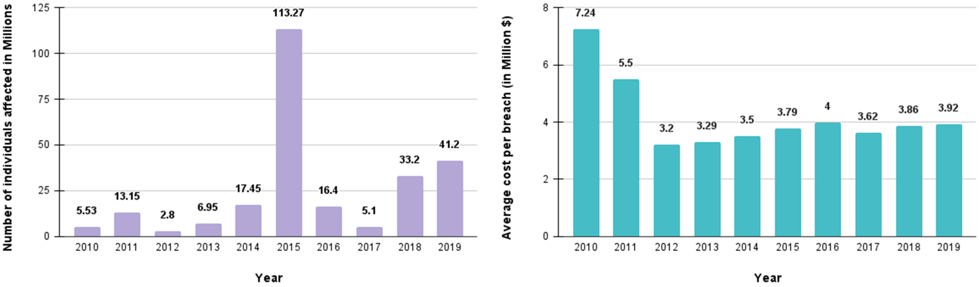

FIGURE 1 Statistical analysis of healthcare data breaches: (a) Number of individuals affected per breach; (b) average cost incurred per breach

them. All the transactions in the blockchain are verified by the process called mining, where all the participating nodes agree to a protocol called consensus. There exist several approaches to mine the transactions. The harder the process of mining is, the harder the transactions are to verify. This makes the network more secure as it is complex to tamper with the mining process. KGs are a powerful tool that can help to incorporate knowledge from multiple sources and thus can be a single source of rich information. The broader objective of SENSIBLE is to deploy a patient-centric digital healthcare ecosystem with the following deliverables:

1. *Secured* and *seamless* integration of fragmented health data;
2. Representing the integrated data with *a single source of modality* with KGs;
3. *Secured consent-based knowledge sharing* amongst the relevant stakeholders;
4. *Personalised healthcare* via precision medicine.

In this paper, we present a private, permissioned blockchain-based data-sharing framework, *SENSIBLE*, to leverage the existing blockchain-based EHR systems. This framework proposes a novel approach of sharing integrated from diverse sources through KGs amongst different stakeholders within the healthcare ecosystem. This research is an initial proof-of-concept for our perspective article discussing the principles behind such an endeavor [19, 20]. With the use of KGs and blockchain, *SENSIBLE* aims to open the potential of transforming the existing healthcare systems towards personalized healthcare and futuristic precision medicine [21] through the following contributions:

1. A secured data sharing framework facilitated through blockchains for disseminating rich information amongst relevant stakeholders while maintaining privacy and security;
2. Representation of integrated data with KGs enabling real-time access to a patient's medical records through smart contracts;

2 | RELATED WORKS

There has been much research using blockchain for sharing data with EHRs [30] and using KGs [31] to extract information from integrated sources. The private [32, 33] and consortium or federated, blockchains [34, 35] have been proposed in the literature, along with their advantages and features in data integrity. For sensitive data applications like health, private, permissioned, or consortium blockchains such as Ethereum, Hyperledger Fabric, Multichain, Corda,

TABLE 1 Feasibility study of using blockchain and knowledge graphs with the relevant works in the last five years.

Work	**Blockchain used/ Type**	**KG**	**Dataset storage**	**Key deliverables**	**Deployed?**
HealthChain [22]	Yes/ Consortium	No	• Off chain (cloud) • On chain (metadata)	• Distributed ledger • privacy • security • high throughput	No
Action-EHR[23]	Yes/ Hyperledger	No	Server/cloud/single machine	Patient-centric EHR sharing for cancer	Yes (US)
Kim et. al[24]	Yes/NA	No	• Off chain (sensitive data) • On chain (encrypted) – less sensitive data	• Reduction of data breach • Hybrid off and on chain storage	No
e-Health[25]	Yes/KSI	No	Centralised national database	• Patient centric centralised system • blockchain (data integrity and access log)	Yes (Estonia)height
Santos et. al[26]	No	Yes	NA	• integration of – biomedical databases – literature – publications • AI/ML algorithms for clinical decision making	open-sourceheight
Rotmensch et. al[27]	No	Yes	NA	• Graph based framework • prediction of diseases from symptoms	No
Hernandez et. al[28]	No	Yes	NA	• knowledge based collaborative framework • GDPR-compliant data handling	Yes (Ireland)
Medicalchain[29]	Yes/ Hyperledger, Ethereum	No	Private servers	• User-focused data sharing of EHR • Dual blockchain networks	Yes (US)
SENSIBLE	Yes/Ethereum	Yes	NA	• Data sharing for enhanced collaboration • Privacy • Security • Knowledge extraction with KGs	Proof-of-Concept (Ireland)

NA, data not available.

etc. are preferred to allow only "trusted entities" for being a part of the blockchain network which ensures maintaining the integrity of the blockchain networks. Table 1 presents a comparative analysis of such works from the past five years. We analyse the blockchain platform used, the use of KGs for extracting information, data storage, and the major objectives of the systems. Most of the works have either used blockchain for EHRs or KGs, but no work has used both for EHRs and data-sharing to the best of our knowledge. All the works such as in [36, 37] have focused on reducing the chances of data tampering while increasing the data security and efficiency of the system through the use of blockchain and cryptographic algorithms. With this objective, the data storage is either private servers or secured clouds, or a combination of off-chain and on-chain storage. Hence, this analysis inspired us to use one of the federated blockchain network with KG to unfold the benefits of both of these technologies can bring into secure and effective collaboration. In the last row of the table, we also propose how *SESINBLE* system differs in comparison with these stateoftheart approaches, which we will discuss in detail in section 3.

3 | METHODOLOGY

This section discusses the architecture and functionalities of the proposed framework.

3.1 | Overview of the *SENSIBLE* Network

SENSIBLE is a decentralized data-sharing framework for stakeholders in the healthcare industry powered by blockchain to maintain the privacy of shared sensitive data. We consider the following major stakeholders of the industry in this research work:- patients, doctors, lab technicians, pharmacists, and insurance providers. The proposed architecture of *SENSIBLE* is illustrated in Figure 2. The overall idea of the framework is to ease the integrated data-sharing across the stakeholders of the industry. For this, we have the following assumption: *The patient data from diverse sources such as hospitals, healthcare providers, wearables, etc. is collected, integrated, processed, compressed, stored, and secured with cryptographic algorithms.* Data processing includes the conversion of data into a standard format, while data compression ensures less storage and high throughput. The process of data collection and processing is out of the scope of this paper. The framework proposes data-sharing once we have the data integrated and available for sharing. The data stored on platforms such as the cloud and servers would have predefined access rights for data access. For instance, patients can predefine the portions of their personal data that can be viewed by which of the stakeholders, giving them consent to edit even in the future. Whenever a stakeholder requests a view of the patient data, this session is secured as a transaction, authenticated, and authorized through smart contracts on the blockchain network. The requested data is extracted from the data storage generating a KG and shared with the stakeholders for the active session duration. The transaction is then verified by the nodes in the networks and appended to the blockchain.

3.2 | Key Components of *SENSIBLE* blockchain network.

We discuss the key entities and components in the blockchain network in the following section.

3.2.1 | Resource-Owner

A resource-owner in the *SENSIBLE* blockchain network is an entity or a person that owns a resource – which is the integrated data in the standard format – and declares the rights on data access. A resource-owner owns a unique

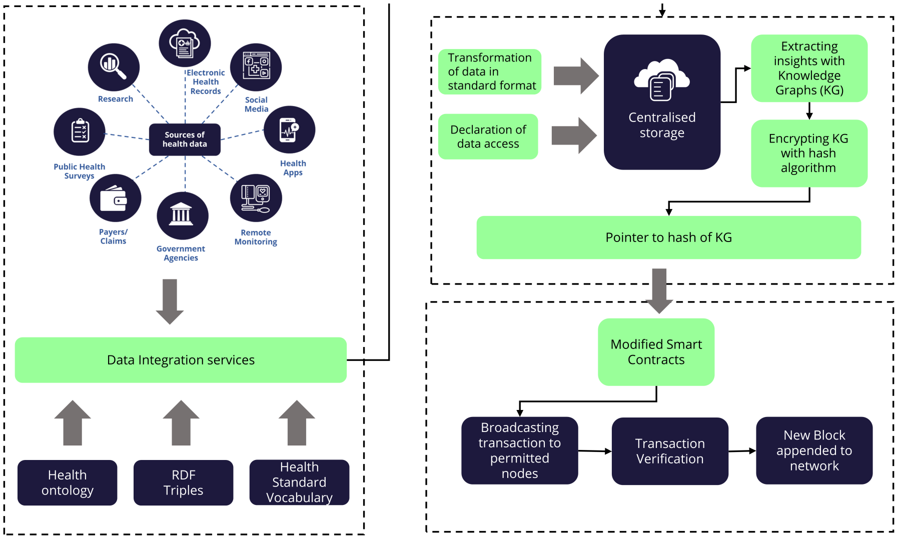
Research
Electronic Health Records
Social Media
Sources of health data
Public Health Surveys
Health Apps
Payers/ Claims
Government Agencies
Remote Monitoring
Data Integration services
Health ontology
RDF Triples
Health Standard Vocabulary
Transformation of data in standard format
Declaration of data access
Centralised storage
Extracting insights with Knowledge Graphs (KG)
Encrypting KG with hash algorithm
Pointer to hash of KG
Modified Smart Contracts
Broadcasting transaction to permitted nodes
Transaction Verification
New Block appended to network

FIGURE 2 Overview of the *SENSIBLE* framework

address in the blockchain network and interacts with the other stakeholders in the network using this unique address.

3.2.2 | Requester

A requester is an entity that requests the resourceowner to access the data for an active session time and can be any of the stakeholders in the healthcare industry, viz. doctors, pharmacists, lab technicians, insurance providers, etc.

3.2.3 | Wallet

A wallet is component holding credentials and transaction details in the blockchain network. In *SENSIBLE*, these transactions take place between a resource owner and the stakeholders of the healthcare industry. The wallet also contains all access permissions of the resources defined by the resource owner for the stakeholders.

3.2.4 | Transactions and Consensus Mechanism

A transaction in blockchain refers to a signed data package that contains a message to send from an externally owned account. The transactions need to follow some rules in order to execute or send across the network. *SENSIBLE* system consists of the following types of transactions on blockchain networks. We discuss them in detail in section 3.3.

Add Patient

Register a new user on the system by verifying that the patient doesn't already exist in the records. A public and private address will be generated upon successful verification of the user. The public address will then be shared with all users, while the private address will only be accessible by the patient via the blockchain wallet. These key pairs are generated with the help of cryptographic algorithms such as Elliptic Curve Cryptography (ECC).

Add patient record

Add an entry to the medical records in the *SENSIBLE* system when a new transaction is updated in the system that will be stored in the respective data storage system.

View patient record

Permit the patient to see his/her own medical records. This transaction first verifies the user with his/her public address, ensuring that only the patient's own medical records are accessible by him/her.

Request patient data

Allow stakeholders (patients, insurance providers, pharmacists) other than patients to request a view of the patient data for viewing. The stakeholders can request the data and can view it when the patient grants access to the stakeholders.

Generate KG

This is a subsequent transaction for the requested patient data which creates a KG for the requested data. When the patient grants the request, the requested data gets extracted from the database and converted to KG.

3.2.5 | Incentive Mechanism

While converting the proposed system to a business model, covering the computational cost of the transactions is equally important. Sensitive applications like healthcare do not intend to create a digital currency from the perspective of sustainability. Hence, a monetaryless incentive mechanism as proposed in [38], or shifting towards federated blockchains with incentive-less consensus algorithm can be a possibility for this in the future.

3.2.6 | Security Analysis

As SESINBLE deals with data-sharing, the possible sensitive points would be during the data-sharing process. Blockchain has the unique characteristic of tracing, which enables it to identify malicious activity on the network. If a malicious user tries to tamper with the KG or session, the account can be traceable from the blockchain network as it will get redirected to the network as illustrated in Figure 3. In *SENSIBLE*, the data is not stored on the blockchain, so the only source of information that blockchain contains is the metadata of transactions. For example, if a malicious user wants to change the contents of Block 3 in Figure 3, the hash of Block 3 will change. As the hash of the block is always the hash of the current blocks and the hash of the previous blocks, all the blocks after block 3 will change, and thus the attack can easily be located. Blockchains record timestamps of each instance of access or modification performed on the patient records, and the cryptographic hash functions make it possible to trace and monitor all activities.

3.3 | Detailed processes of *SENSIBLE* Framework

We discuss the working of the major transactions of the *SENSIBLE* network in detail which were presented in section 3.

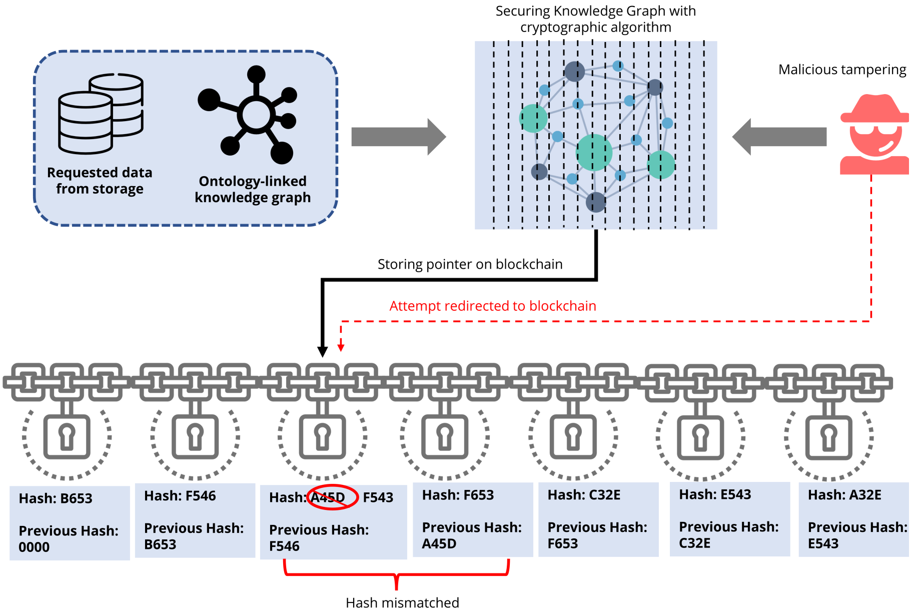

FIGURE 3 Example of data tampering prevention in *SENSIBLE* network.

3.3.1 | User Creation

Figure 4 shows the process of creating users on the *SENSIBLE* network. A session gets initiated when a new user is requested to register on the network. The user is verified whether he/she/they exist in the system and upon successful verification, he/she/they is registered in the network with a predefined role from any of the following - patient, doctor, pharmacist, insurance provider, etc. The user is added to the network with the generation of public and private key pairs and the transaction is verified and appended to the blockchain. The unique identifiers for users are significant because roles overlap in real-world scenarios, as any user on the network can have multiple roles such as patient, as well as concurrently be a doctor or a pharmacist. Once the user is registered on the network, he/she/they can define the access rights for all the resources owned by him/her/them.

3.4 | Data Sharing with access control model

One significant component of the *SENSIBLE* system is the access control model which defines the complete ownership of the data to the resource-owner or the patient and its access by other stakeholders in healthcare. The patient can define portions of their health data accessible to different stakeholders over the network, including any or all of their entire historical data on the network. The patient also has the right to modify the access rights of any stakeholder at any instance of time. This makes the system user-driven and secured while ensuring data privacy. Figure 5 shows the process of data sharing across stakeholders with KGs extracting insights from the diverse integrated data sources. When a stakeholder requests data, smart contracts authenticate him/her/them with the public address on the network and verify that they have the right to view the requested data. A consent request is sent to the patient upon successful verification. Once the patient approves the request to view the data, the requested data is then extracted and linked

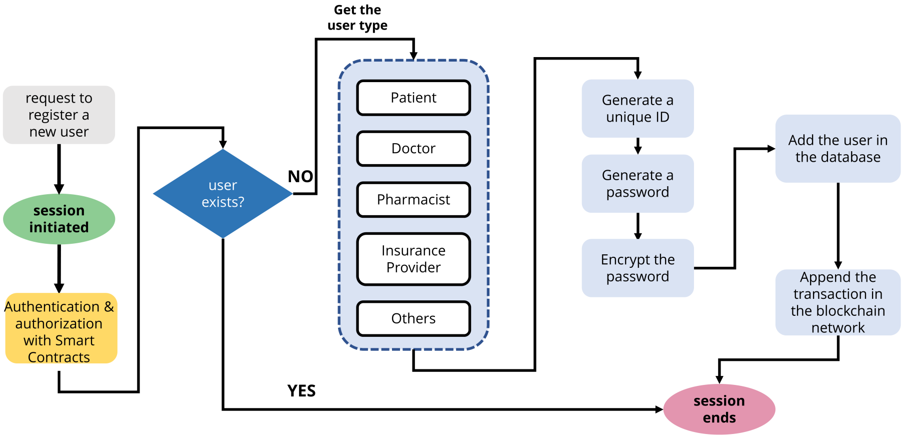

FIGURE 4 Flowchart for the process of registering a new user in the *SENSIBLE* network.

with ontologies to develop a relationship with them to generate a KG. The tuples form the nodes of the KGs while the ontologies form the edges. The KGs are dynamically generated upon data request and the updated details are stored in the data source again after the session ends. The transactions are then verified within the peer-to-peer network of the blockchain and appended to it. We don't store any KGs using blockchain, as they are dynamically created during the session. However, we update the data source with the encounter details. The advantage of performing all data sharing activities through blockchain is the mitigation of the risk of highly sensitive data being compromised by some malicious user. Data Integration from different modalities leads to a data sharing Framework.

4 | PROOF OF CONCEPT OF *SENSIBLE* WITH ETHEREUM

In this section, we present a pilot study of our proposed *SENSIBLE* network and discuss the dataset used for experimentation, network setup and its performance.

4.1 | Dataset Details

For *SENSIBLE* pilot testing, we have used the samples of synthetic health records generated with open-source software, Synthea [39]. The dataset consists of 14 different files of data covering healthcare transactions in the healthcare industry for synthetic patients. The 10 most common encounters in the healthcare industry with the 10 highest morbidity chronic conditions have been provided in the dataset. The data is based on the statistics of the United States and also adheres to universal healthcare standards [40, 41] such as Health Level – 7 (HL7) and, in multiple formats like CSV, FHIR, and C-CDA. The data of 1000 sample patients in CSV format were used for the *SENSIBLE* network testing. Figure 6 shows the database structure - the entities, their properties, and relations amongst them. This illustrates how complex the ontologies get when multimodal data sources and stakeholders are involved in the system.

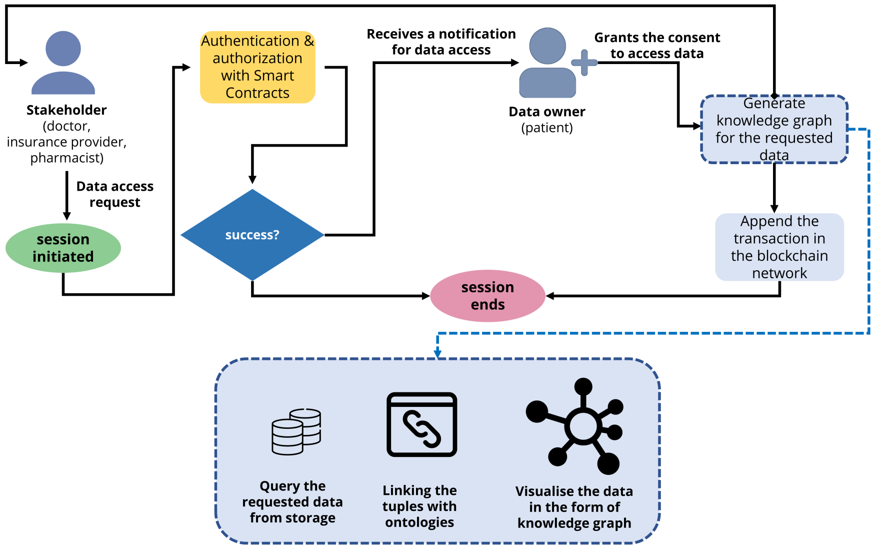
Stakeholder
(doctor,
insurance provider,
pharmacist)
Data access request
session initiated
Authentication & authorization with Smart Contracts
success?
Receives a notification for data access
Data owner
(patient)
Grants the consent to access data
Generate knowledge graph for the requested data
Append the transaction in the blockchain network
session ends
Query the requested data from storage
Linking the tuples with ontologies
Visualise the data in the form of knowledge graph

FIGURE 5 Illustration of data sharing with knowledge graph using the access control mechanism

4.2 | Private Blockchain Network

We used the Ethereum blockchain for our pilot study because of its characteristics like its permissioned nature, scalability, efficiency, and finality. We set up a private 7 remotes node Ethereum network with the following configurations: {*Node 1*: 8 GB RAM, 1.19 GHz, *Node 2*: 4 GB RAM, 1.80 GHz, *Node 3*: 8 GB RAM, 1.19 GHz, *Node 4*: 8 GB RAM, 2.5 GHz, *Node 5*: 8 GB RAM, 2.4 GHz, *Node 6*: 12 GB RAM, 2.20 GHz, *Node 7*: 16 GB RAM, 2.20 GHZ}

As it was a permissioned network, the root node that defines the mining rights to other nodes was at Node 1, and each of Node 2, Node 3, and Node 4 were given rights to mine the transactions. Nodes 5 through 7 were not given permission for mining. This setup was to accurately model a blockchain network in which all nodes do not have the same mining permissions. The blockchain private network was built using geth and Puppet tool and Web3j library for smart contracts. Geth (Go Ethereum) [42] is a command-line client interface tool that allows users to create and interact with private Ethereum blockchains. Puppeth [43] is a command line interface tool that helps generate genesis blocks, which are initialization files for the nodes to get registered on the network. Web3j [44] is a library used to invoke smart contracts. All the remote nodes were connected to a network using a virtual private network. After establishing a connection between the nodes, to test and verify the connection between them, we start mining on the first node and verify that the changes are reflected on the other nodes. The proof-of-Work consensus was used by the nodes to verify and mine the transactions.

4.3 | Stakeholder Co-operation and Communication with Knowledge Graph

The Synthea dataset and the private blockchain network were used to test the SENSIBLE framework. To illustrate how data sharing through KGs and blockchain works, let us consider a scenario when a doctor, pharmacist, and insurance provider requests the medical history of the patient. For this scenario, we have used a sample of a patient from the dataset with patient id “76982e06-f8b8-4509-9ca3-65a99c8650fe” and the interactions amongst the stakeholders

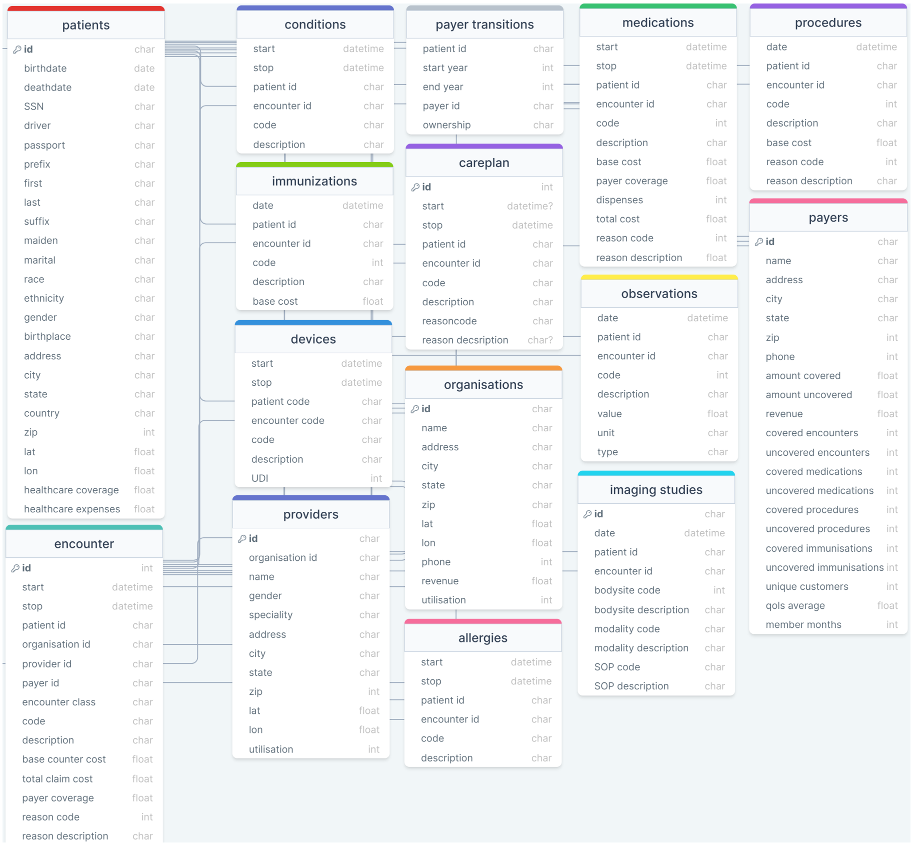

FIGURE 6 Illustration of integrated data of patients across multiple files and their interdependencies in Synthea dataset.

are as the following:

1. When the doctor requests the data, a smart contract would get initiated on the blockchain and will verify the authentication of the doctor with his public address, along with his authorization to view the data. After successful verification, the data would then be converted to a KG. An illustration of a KG generated for this patient using the data from all 14 CSV files is shown in Figure 7. The nodes in the KG represent the attributes of the patient history while the ontologies on the edges are used to develop a logical relation between them. In this pilot study, we have pre-defined the ontologies for the dataset and presented them in Table, although the automatic generation of ontologies will be in future work. After the successful execution of a transaction, the transaction is verified on the blockchain network and appended to the chain. ;
2. When a pharmacist wants to give the prescribed medication to the patient, he will only see the relevant portions from the complex KG;
3. Similarly, when an insurance provider wants to refer to the patient data for insurance claims, they can only see the relevant data from the patient history.

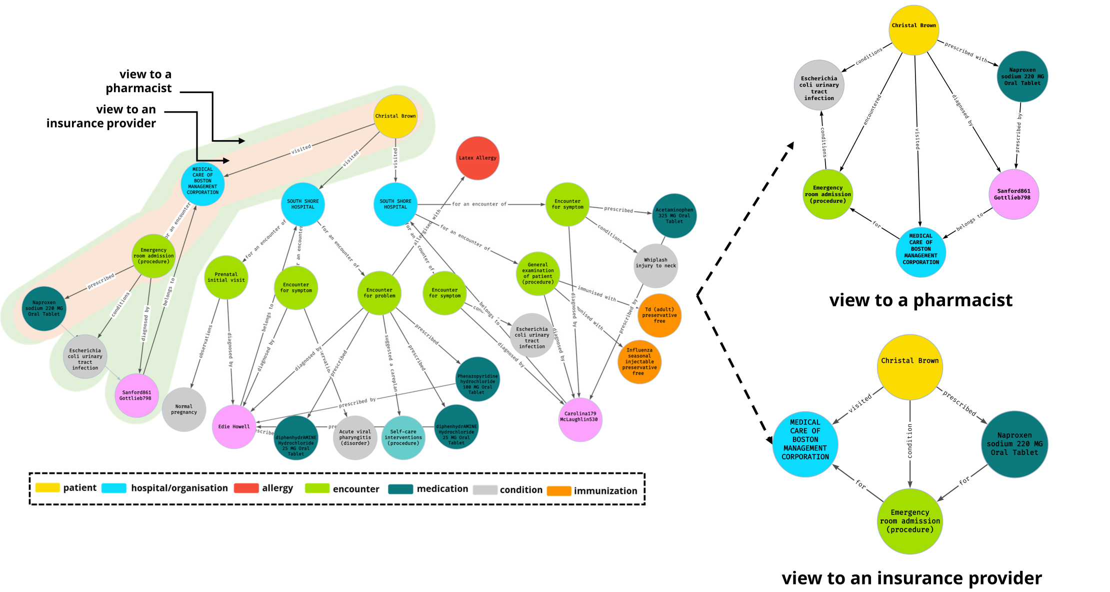
view to a pharmacist
view to an insurance provider
Christal Brown
MEDICAL CARE OF BOSTON MANAGEMENT CORPORATION
SOUTH SHORE HOSPITAL
SOUTH SHORE HOSPITAL
Latex Allergy
Emergency room admission (procedure)
Naproxen sodium 220 MG Oral Tablet
Escherichia coli urinary tract infection
Sanford861 Gottlieb798
Prenatal initial visit
Encounter for symptom
Encounter for problem
Encounter for symptom
Encounter for symptom
General examination of patient (procedure)
Acetaminophen 325 MG Oral Tablet
Whiplash injury to neck
Td (adult) preservative free
Influenza seasonal injectable preservative free
Escherichia coli urinary tract infection
Phenazopyridine hydrochloride 100 MG Oral Tablet
diphenhydrAMINE Hydrochloride 25 MG Oral Tablet
diphenhydrAMINE Hydrochloride 25 MG Oral Tablet
Normal pregnancy
Edie Howell
Acute viral pharyngitis (disorder)
Self-care interventions (procedure)
Carolina179 McLaughlin530
visited
for an encounter of
prescribed
conditions
diagnosed by
belongs to
observations
allergised with
suggested a careplan
immunised with
prescribed by
patient
hospital/organisation
allergy
encounter
medication
condition
immunization
Christal Brown
Escherichia coli urinary tract infection
Naproxen sodium 220 MG Oral Tablet
Emergency room admission (procedure)
Sanford861 Gottlieb798
MEDICAL CARE OF BOSTON MANAGEMENT CORPORATION
conditions
prescribed with
encountered
visited
diagnosed by
prescribed by
for
belongs to
view to a pharmacist
Christal Brown
MEDICAL CARE OF BOSTON MANAGEMENT CORPORATION
Naproxen sodium 220 MG Oral Tablet
Emergency room admission (procedure)
visited
prescribed
condition
for
for
view to an insurance provider

FIGURE 7 Data sharing among a single patient and three different stakeholders in the *SENSIBLE* network. The KGs are generated using patient detail from the dataset. The three different views of the KG are for three different transactions amongst stakeholders: a doctor and a patient; (b) a pharmacist and a patient; (c) an insurance provider and a patient. The knowledge shared differs according to the predefined rights of the stakeholders to access the relevant portions of the data and the data requested at the time of the transaction. The respective colors represent different attributes of the patient medical history.

An illustrated example of all these scenarios is depicted in Figure 7. Such information-rich customized KGs hold the potential to disrupt future healthcare ecosystems by shifting towards personalized healthcare tailored according to each individual leading to improved health, an effective source of knowledge for the healthcare providers, and a crucial step toward precision medicine.

5 | PERFORMANCE ANALYSIS OF SENSIBLE

For blockchain network performance, time, memory usage, and cost are significant factors for measuring the scalability and efficiency of the network. We have analyzed the performance of the network using these factors.

5.1 | Time Analysis

For analyzing the performance of the blockchain network in terms of time, we have considered the execution time, latency, and throughout the system which is explained in detail as follows.

5.1.1 | Execution Time

The time taken for the transactions to be submitted on the blockchain network and verified by the nodes is called the execution time. There were 16 different types of transactions (including all type of authorization, authentication, and access control mechanism) in the *SENSIBLE* network. For this analysis, only two nodes were allowed for mining (Node 1 and Node 2) to test the performance on a small network. Initially, we conducted an experiment to test the support of the *SENSIBLE* network for simultaneous access of the data for the basic transaction of creating users on the network. The average time taken by each of the transactions over 3 rounds of transactions is shown in Table 2. As observed from the table, the time taken by the complex transactions like registering a new user or allowing consent from the patient takes more time compared to other ones. When these transactions are performed and submitted simultaneously, the average time for a single transaction is around 15 seconds as illustrated in Figure 8.

TABLE 2 Time analysis for different types of transactions.

Type of transaction	Time (s)
setPatientData	30.62
getpatientData	0.215667
setDoctor	35.331
requestAccess	27.17067
getMyPendingRequests	0.263333
allowAccess	20.13833
accessDataByDoctor	0.207333
closeAccess	20.18033
getTransactionHistory	0.209
addPatientHash	22.7315
getPatientHash	0.307667
setPharmacists	20.13733
requestAccess	20.11967
allowAccess	20.154
closeAccess	20.155
getPatientHash	0.124333

As the number of transactions increased, the average time taken by the system for a batch of 100 transactions was 255 seconds. This gives an idea of scalability during simultaneous submission of transactions as total time would increase when the number of transactions is submitted simultaneously on the system.

5.1.2 | Latency

The time difference between submission of a transaction and confirmation of a transaction is called *latency*. The less the value of latency, the more secure and efficient the blockchain network is. Considering the abovementioned

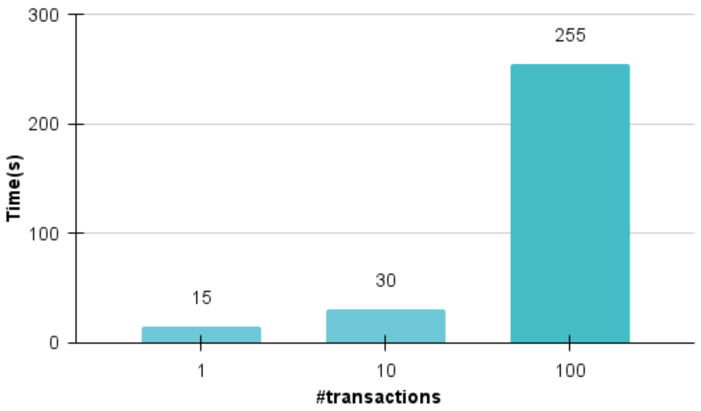

FIGURE 8 Time analysis of *SENSIBLE* blockchain network for a batch of 1, 10, and 100 transactions.

batch of 100 transactions, the average latency of the *SENSIBLE* network is 2.5 seconds. The value of latency shows a promising nature within 2 nodes of the network which can be scaled in the future for real-world applications.

5.1.3 | Throughput

The number of successfully confirmed transactions in a second is termed *throughput*. The more the throughput, the more efficient the blockchain network is. With a batch of 100 transactions, the throughput of *SENSIBLE* is 0.04 transactions per second tps. This is because some of the transactions from these are complex including of their authentications.

5.2 | Cost Analysis

We also tested the efficiency of the proposed framework in terms of computational efforts and their associated cost. The computational cost to verify the transactions is paid by the user who requests the transaction verification, in the form of transaction fees [45]. The transaction fees are dependent on multiple factors, such as the complexity of the smart contracts, the gas price paid, and the frequency of smart contracts used in the transactions. One significant factor impacting transaction fees is the gas limit on transaction costs. The gas limit determines the maximum gas price a block can have and the number of transactions that a block can contain. The transaction fee is calculated using Equation 1.

$$transaction\ fee = gas\ consumed * gas\ price \tag{1}$$

Here, gas measures the computational efforts required to execute any operations on the Ethereum blockchain network and is measured in the units *gwei*. The computational efforts can also be calculated with Ethereum's own cryptocurrency, *Ether*, with a single ether currently worth approximately $1,679.90 as of 25th August 2022 [46].

In our pilot study, we calculated the cost of transactions on a public network of ganache-cli [47], which gives

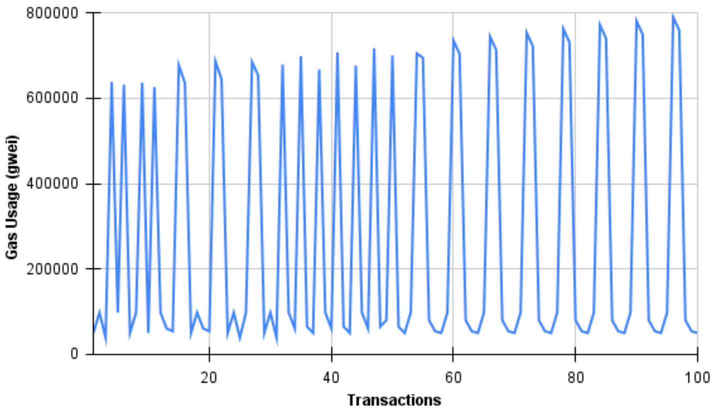

FIGURE 9 Analysis of transaction fees for verification of transactions on *SENSIBLE* network for a batch of 100 transactions on the ganache-cli platform.

direct cost of gas used for verification of transactions, for a batch of 100 transactions, and the costs of transactions are shown in Figure 9.

The parameters while performing the transactions were: *gas limit: 6721975*, *gas price: 20 gwei*. We performed 100 transactions over the *SENSIBLE* network and calculated the transaction costs to verify the transactions in the network. The average gas usage for 100 transactions was 276940.5941 *gwei*. Hence, with equation 1, the transaction fee would be:

$$transaction\ fee = 276940.5941 * 20$$

$$= 5538811.882\ gwei.$$

This transaction fee of 5538811.882 *gwei* would be equivalent to 0.008424 ETHER or 14.15 USD. This analysis gives an idea about identifying the impact of various factors on transaction costs while planning a business model.

5.3 | Local produced blocks amount

As discussed in the earlier section, the gas limit determines the number of transactions that can be contained in a block. However, the configurations of nodes also have a significant impact on the mining of transactions and confirmation. To analyze the impact of the performance of the network with respect to node configurations, we carried out an experiment by giving different nodes permission to mine with different Random Access Memory (RAM) sizes. Among the nodes with different configurations, we present a comparison of the number of blocks mined when a batch of 100 transactions was run on the system Table 3. It can be observed that the maximum nodes of nodes were mined as 34 and by the node with maximum RAM amongst the three, i.e. 16 GB, nearly 5x times the basic node. Blocks mined have an impact indirectly on latency and throughput of the network. Hence, we need to ensure at least 8 GB of RAM

TABLE 3 Effect of RAM size on transactions mined.

RAM size (GB)	#transactions mined
4	7
8	13
16	34

TABLE 4 Performance analysis of the system against network size.

#transactions	Time (s)			Latency (tps)			Throughput		
	2	3	4	2	3	4	2	3	4
1	15	19	22	15	19	22	0.067	0.08	1.87
10	30	51	56	3	5.1	5.6	0.33	0.87	2.85
100	255	307	345	2.55	3.07	3.45	1.44	2.35	4.85

tps: transactions per second

when any node is registered on the network for mining.

5.4 | Peer-to-peer network analysis

We also tested the scalability of the network by varying the network size with 2, 3, and 4 nodes for mining the transactions and captured the results in Table 4. As the network size increases with the number of nodes, there is also an increase in performance and trust as more nodes take part in the consensus [48]. However, if the node has a lower configuration such as 4GB RAM, there is a trade-off in transaction confirmation time. We observed an increase in execution time with an increase in the number of nodes with an effect of node configuration. Hence, the configuration of the system serving as a node shall be powerful enough to reduce the confirmation time and the chances to tamper and fork the chain.

5.5 | System Failure Analysis

The KGs are dynamically created in the blockchain network and hence not stored o the chain, which ensures that the sensitive data has no chance of being tampered with on the chain. However, as we have used the Proof-of-Work consensus algorithm with the Ethereum network, there is a single point of failure of the system when all of the nodes stop working at once. This can be prevented in the future by shifting towards federated blockchain platforms such as Hyperledger by using the Byzantine Fault Tolerant protocol as the consensus algorithm.

6 | CONCLUSIONS AND FUTURE WORK

In this research, we demonstrated *SENSIBLE*, a data-sharing framework using a blockchain network that includes following key aspects:

1. Compared with traditional data integration solutions, our approach adopted cutting-edge semantic technologies such as KGs to manage the complexity of massive data-sharing transactions;
2. Using the Ethereum blockchain technology, our framework offers secured data access during multiple data-sharing transactions amongst the stakeholders;

Through these benefits, the proposed framework can tackle the core concerns of current smart healthcare applications and provide the stakeholders with a simple, secure, and convenient data-sharing platform. In future work, we plan to design Graph Neural Networks for recommendations in personalized treatments by feeding the GNNs with KGs from the SENSIBLE framework. The application of our framework can help users to share sensitive data in these healthcare domains securely and achieve a much better and comprehensive understanding of the complicated scenarios which can lead to reduced medical errors and thus potentially save lives.

ACKNOWLEDGMENTS

This work was supported by the Science Foundation Ireland Grant #16/IA/4605. This work was conducted with the financial support of the Science Foundation Ireland Centre for Research Training in Artificial Intelligence under Grant No. 18/CRT/6223.

CONFLICT OF INTEREST STATEMENT

The authors have no conflict of interest relevant to this article.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available at https://doi.org/10.1093/jamia/ocx079.

references

[1] Yanamadala S, Morrison D, Curtin C, McDonald K, Hernandez-Boussard T. Electronic health records and quality of care an observational study modeling impact on mortality, readmissions, and complications. Medicine (United States) 2016;95. https://journals.lww.com/md-journal/Fulltext/2016/05100/Electronic_Health_Records_and_Quality_of_Care__An.10.aspx.

[2] Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. Journal of the American Medical Informatics Association : JAMIA 2013;20:117–121. https://pubmed.ncbi.nlm.nih.gov/22955496/.

[3] Caine K, Hanania R. Patients want granular privacy control over health information in electronic medical records. Journal of the American Medical Informatics Association 2013 1;20:7–15. https://academic.oup.com/jamia/article/20/1/7/730602.

[4] Wu LT, Brady KT, Spratt SE, Dunham AA, Heidenfelder B, Batch BC, et al. Using electronic health record data for substance use Screening, Brief Intervention, and Referral to Treatment among adults with type 2 diabetes: Design of a National Drug Abuse Treatment Clinical Trials Network study. Contemporary clinical trials 2016 1;46:30–38. https://pubmed.ncbi.nlm.nih.gov/26563446/.

[5] Tan MH, Bernstein SJ, Gendler S, Hanauer D, Herman WH. Design, development and deployment of a Diabetes Research Registry to facilitate recruitment in clinical research. Contemporary clinical trials 2016 3;47:202–208. https://pubmed.ncbi.nlm.nih.gov/26825022/.

[6] Hammack-Aviran CM, Brelsford KM, McKenna KC, Graham RD, Lampron ZM, Beskow LM. Research Use of Electronic Health Records: Patients' Views on Alternative Approaches to Permission. AJOB Empirical Bioethics 2020 7;11:172–186. https://www.tandfonline.com/doi/abs/10.1080/23294515.2020.1755383.

[7] Are your home and your belongings internet-connected? - Products Eurostat News - Eurostat;. Accessed: 2022-04-24. https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210225-1.

[8] Big data analytics in healthcare: promise and potential. Health Information Science and Systems 2014 2;2. /pmc/articles/PMC4341817//pmc/articles/PMC4341817/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/.

[9] Fernandes LM, O'Connor M, Weaver V. Big Data, Bigger Outcomes. Journal of AHIMA 2012;83:38–43.

[10] Nagori M, Patil A, Deshmukh S, Vaidya G, Rahangdale M, Kulkarni C, et al. Mutichain Enabled EHR Management System and Predictive Analytics. Smart Innovation, Systems and Technologies 2020;165:179–187.

[11] Holmes J, Beinlich J, Boland MR, Bowles K, Chen Y, Cook T, et al. Why Is the Electronic Health Record So Challenging for Research and Clinical Care? Methods of information in medicine 2021 07;60.

[12] Bindra P, Kshirsagar M, Ryan C, Vaidya G, Gupt KK, Kshirsagar V. Insights into the Advancements of Artificial Intelligence and Machine Learning, the Present State of Art, and Future Prospects: Seven Decades of Digital Revolution. In: Satapathy SC, Bhateja V, Favorskaya MN, Adilakshmi T, editors. Smart Computing Techniques and Applications Singapore: Springer Singapore; 2021. p. 609–621.

[13] Abouelmehdi K, Beni-Hessane A, Khaloufi H. Big healthcare data: preserving security and privacy. Journal of Big Data 2018 12;5:1–18. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-017-0110-7.

[14] Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, et al. Healthcare Data Breaches: Insights and Implications. Healthcare 2020 6;8. /pmc/articles/PMC7349636//pmc/articles/PMC7349636/?report=abstracthttps://www.ncbi.nlm.nih.gov/pmc/articles/PMC7349636/.

[15] 2018 Data Breach Investigations Report.;. Accessed: 2022-04-24. https://www.verizon.com/business/resources/reports/.

[16] December 2019 Healthcare Data Breach Report;. Accessed: 2022-04-24. https://www.hipaajournal.com/december-2019-healthcare-data-breach-report/.

[17] Nakamoto S, Bitcoin: A Peer-to-Peer Electronic Cash System; 2008. https://bitcoin.org/bitcoin.pdf.

[18] Fensel D, Şimşek U, Angele K, Huaman E, Kärle E, Panasiuk O, et al. In: Introduction: What Is a Knowledge Graph? Cham: Springer International Publishing; 2020. p. 1–10. https://doi.org/10.1007/978-3-030-37439-6_1.

[19] Yao Y, Kshirsagar M, Vaidya G, Ducrée J, Ryan C. Convergence of Blockchain, Autonomous Agents, and Knowledge Graph to Share Electronic Health Records. Frontiers in Blockchain 2021 4;4.

[20] Yao Y, Kshirsagar M, Vaidya G, Ryan C. Using a bio-inspired model to facilitate the ecosystem of data sharing in smart healthcare. Evo* 2021 – Late-Breaking Abstracts Volume 2021;p. 25–30.

[21] Vicente AM, Ballensiefen W, Jönsson JI. How personalised medicine will transform healthcare by 2030: The ICPerMed vision. Journal of Translational Medicine 2020 4;18:1–4. https://translational-medicine.biomedcentral.com/articles/10.1186/s12967-020-02316-w.

[22] Chenthara S, Ahmed K, Wang H, Whittaker F, Chen Z. Healthchain: A novel framework on privacy preservation of electronic health records using blockchain technology. PLOS ONE 2020 12;15.

[23] Dubovitskaya A, Baig F, Xu Z, Shukla R, Zambani PS, Swaminathan A, et al. ACTION-EHR: Patient-Centric Blockchain-Based Electronic Health Record Data Management for Cancer Care. J Med Internet Res 2020;22(8):e13598 https://wwwjmirorg/2020/8/e13598 2020 8;22:e13598. https://www.jmir.org/2020/8/e13598.

[24] Kim TM, Lee SJ, Chang DJ, Koo J, Kim T, Yoon KH, et al. DynamiChain: Development of Medical Blockchain Ecosystem Based on Dynamic Consent System. Applied Sciences 2021;11(4). https://www.mdpi.com/2076-3417/11/4/1612.

[25] Mata IDL, Kluge H, Kluge H, Monteiro A, Monteiro A, Odone A, et al. Health in the Digital Society: the experience of Estonia. European Journal of Public Health 2018 11;28. https://academic.oup.com/eurpub/article/28/suppl_4/cky213.014b/5191946.

[26] A knowledge graph to interpret clinical proteomics data. Nature Biotechnology 2022;https://doi.org/10.1038/s41587-021-01145-6.

[27] Rotmensch M, Halpern Y, Tlimat A, Horng S, Sontag D. Learning a Health Knowledge Graph from Electronic Medical Records. Scientific Reports 2017 12;7.

[28] Hernandez J, McKenna L, Brennan R. TIKD: A Trusted Integrated Knowledge Dataspace For Sensitive Healthcare Data Sharing. In: 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC); 2021. p. 1855–1860.

[29] Medicalchain - Whitepaper 2.1;. Accessed: 2022-08-25. https://medicalchain.com/Medicalchain-Whitepaper-EN.pdf.

[30] Qazi M, Kulkarni D, Nagori M. Proof of Authenticity-Based Electronic Medical Records Storage on Blockchain. Smart Innovation, Systems and Technologies 2020;165:297–306.

[31] Mirza A, Nagori M, Kshirsagar V. Constructing Knowledge Graph by Extracting Correlations from Wikipedia Corpus for Optimizing Web Information Retrieval. IEEE; 2018. .

[32] Yue X, Wang H, Jin D, Li M, Jiang W. Healthcare Data Gateways: Found Healthcare Intelligence on Blockchain with Novel Privacy Risk Control. Journal of medical systems 2016 10;40. https://pubmed.ncbi.nlm.nih.gov/27565509/.

[33] Castaldo L, Cinque V. Blockchain-based logging for the cross-border exchange of ehealth data in Europe. Communications in Computer and Information Science 2018;821:46–56. https://link.springer.com/chapter/10.1007/978-3-319-95189-8_5.

[34] Xu B, Xu LD, Wang Y, Cai H. A distributed dynamic authorisation method for Internet+ medical healthcare data access based on consortium blockchain. https://doiorg/101080/1751757520211922757 2021;https://www.tandfonline.com/doi/abs/10.1080/17517575.2021.1922757.

[35] Zhang A, Lin X. Towards Secure and Privacy-Preserving Data Sharing in e-Health Systems via Consortium Blockchain. Journal of Medical Systems 2018 8;42:1–18. https://link.springer.com/article/10.1007/s10916-018-0995-5.

[36] Ryan C, Kshirsagar M, Vaidya G, Cunningham A, Sivaraman R. Design of a cryptographically secure pseudo random number generator with grammatical evolution. Scientific Reports 2022;12(1).

[37] Chen W, Zhu S, Li J, Wu J, Chen CL, Deng YY. Authorized Shared Electronic Medical Record System with Proxy Re-Encryption and Blockchain Technology. Sensors 2021;21(22). https://www.mdpi.com/1424-8220/21/22/7765.

[38] Daraghmi EY, Daraghmi YA, Yuan SM. MedChain: A Design of Blockchain-Based System for Medical Records Access and Permissions Management. IEEE Access 2019;7.

[39] Walonoski J, Kramer M, Nichols J, Quina A, Moesel C, Hall D, et al. Synthea: An approach, method, and software mechanism for generating synthetic patients and the synthetic electronic health care record. Journal of the American Medical Informatics Association 2018 3;25.

[40] Hoofnagle CJ, van der Sloot B, Borgesius FZ. The European Union general data protection regulation: what it is and what it means. Information Communications Technology Law 2019 1;28.

[41] Walinjkar A. FHIR Tools for Healthcare Interoperability. Biomedical Journal of Scientific Technical Research 2018 10;9.

[42] Go Ethereum;. Accessed: 2022-04-24. https://geth.ethereum.org/.

[43] Puppeth;. Accessed: 2022-04-24. https://github.com/puppeth.

[44] web3.js - Ethereum JavaScript API — web3.js 1.0.0 documentation;. Accessed: 2022-04-24. https://web3js.readthedocs.io/en/v1.7.1/.

[45] Laurent A, Brotcorne L, Fortz B. Transactions fees optimization in the Ethereum blockchain. Blockchain: Research and Applications 2022 3;p. 100074.

[46] Cryptocurrency prices, charts and market capitalizations. CoinMarketCap;. Accessed: 2022-04-24. https://coinmarketcap.com/.CoinMarketCap.

[47] ganache-cli - npm;. Accessed: 2022-04-24. https://www.npmjs.com/package/ganache-cli.

[48] Pongnumkul S, Siripanpornchana C, Thajchayapong S. Performance Analysis of Private Blockchain Platforms in Varying Workloads. IEEE; 2017. .