Graduate Degree Program in

<u>Applied Computing</u>

Academic Doctorate

Alex Roehrs

OmniPHR: A Blockchain based Interoperable Architecture for Personal Health Records

São Leopoldo, 2019

Alex Roehrs

OMNIPHR: A BLOCKCHAIN BASED INTEROPERABLE ARCHITECTURE FOR PERSONAL HEALTH RECORDS

Dissertation presented as a partial requirement to obtain the Doctor's degree from the Applied Computing Graduate Program of the Universidade do Vale do Rio dos Sinos - UNISINOS

Advisor:
Prof. Dr. Cristiano André da Costa

Co-advisor:
Prof. Dr. Rodrigo da Rosa Righi

São Leopoldo
2019

R713o Roehrs, Alex.

OmniPHR : a Blockchain based interoperable architecture for personal health records / Alex Roehrs. – 2019.

139 f. : il. ; 30 cm.

Tese (doutorado) – Universidade do Vale do Rio dos Sinos, Programa de Pós-Graduação em Computação Aplicada, 2019.

"Advisor: Prof. Dr. Cristiano André da Costa Co-advisor: Prof. Dr. Rodrigo da Rosa Righi."

1. Registro de saúde. 2. Blockchain. 3. Interoperabilidade semântica. 4. Processamento de linguagem natural (Computação). 5. Sistemas distribuídos. I. Título.

CDU 004

Dados Internacionais de Catalogação na Publicação (CIP)
(Bibliotecária: Amanda Schuster – CRB 10/2517)

Alex Roehrs

OmniPHR: A Blockchain based Interoperable Architecture for Personal Health Records

Dissertation presented to the Universidade do Vale do Rio dos Sinos – Unisinos, as a partial requirement to obtain the title of Doctor in Applied Computing.

Approved on 08/16/2019.

EXAMINATION BOARD

Prof. Hyggo Oliveira de Almeida, Ph.D. – Universidade Federal de Campina Grande (UFCG)

Prof. Claudio Fernando Resin Geyer, Ph.D. – Universidade Federal do Rio Grande do Sul (UFRGS)

Prof. Sandro José Rigo, Ph.D. – Universidade do Vale do Rio dos Sinos (UNISINOS)

Prof. Cristiano André da Costa, Ph.D. (Advisor)

Prof. Rodrigo da Rosa Righi, Ph.D. (Co-advisor)

Viewed and allowed to print.
São Leopoldo, August 26, 2019.

Prof. Rodrigo da Rosa Righi, Ph.D.
Chair of the Applied Computing Graduate Program

To my parents, Flórida and Gervásio (in memoriam).
To my loves, Eline, Nicole and Max.

ACKNOWLEDGMENTS

I would like to start by thanking my advisor, Professor Cristiano André da Costa, who accepted me as his student's guidance from the specialization, master's and doctorate. I am very grateful for his mentorship and support throughout this period.

I would also like to thank the support of my co-advisor and course coordinator, Professor Rodrigo da Rosa Righi. Furthermore, I appreciate all the excellent professors and staff of Unisinos PPGCA (Postgraduate Program in Applied Computing), especially the professors Sandro Rigo, Kleinner Farias de Oliveira, and Jorge Victoria Barbosa. Likewise, I would like to thank the project colleagues who helped me in the development of prototypes and experiments, in particular, Matheus Wichman, André H. Mayer, and Jorge Teixeira. I would also like to thank all of the colleagues from the uHospital project, led by Professor Cristiano, who in one way or another collaborated a lot to carry out this work.

Special thank to my beloved spouse Eline, as well my daughter Nicole, and to my son Max who is coming. You were my inspiration, and I have much to thank you for your unconditional support. Thanks to my family, especially my parents, Florida and Gervásio (in memoriam) for the incentives and teachings.

I want to express a special thanks to Unisinos, especially to coordinating Professors Denise Bandeira, Margrit Krug, Luciano Ignaczak and Mateus Raeder, as well as ASAV (Associação Antônio Vieira), in particular, Marcos Andre Knewitz and Valter Pavoni, for their institutional support.

Finally, thanks to the Hospital de Clinicas de Porto Alegre, in particular, Valter Ferreira da Silva and Professor José Roberto Goldim, Professor Hyggo Almeida from UFCG, as well as Professor Douglas C. Schmidt of Vanderbilt University, for their support.

CAPES / CNPq

The author would like to thank the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) and Brazilian National Council for Scientific and Technological Development (CNPq) for supporting this work.

ABSTRACT

CONTEXT: The advances in the Health Information Technology (HIT) brought many benefits to the health care area, especially to the digital storage of patients' health records. However, it is still a challenge to have a unified viewpoint of patients' health history, because typically, health data is scattered among different health organizations. Furthermore, there are several standards for these records, some of them open and others proprietary. Usually, health records are stored in databases within health organizations and generally do not have external access. This situation applies mainly to cases where health care providers maintain patients' data, known as EHR (Electronic Health Record). In the case of PHR (Personal Health Record), in which patients by definition can manage their health records, they usually have no control over their data stored in health care providers' databases. Even with adopted standards, patients often need to explain over and over their health information when they are taken care at different locations. This problem hinders the adoption of PHR. **OBJECTIVE:** Thereby, we envision two main challenges regarding PHR context: first, how patients could have a unified view of their scattered health records, and second, how health care providers can access up-to-date data regarding their patients, even though changes occurred elsewhere. The scientific contribution is to propose an architectural model based on Blockchain to support a distributed PHR, where patients can maintain their health history in a unified viewpoint, from any device anywhere. Likewise, the scientific contribution for health care providers seeks to promote the possibility of having their patients' data interconnected among health organizations. **METHOD:** The methodology consists in proposing and prototyping an application model named OmniPHR ('Omni' comes from omnipresent) as a distributed model to integrate PHRs. The method to evaluate the model includes assessing the network performance, interoperability, and semantic integration of different health standards, using a real database from anonymized patients. **RESULTS:** The evaluations demonstrate the feasibility of the model in maintaining health records distributed in an architecture model that promotes a unified view of PHR with the scalability of the solution. As a result, we evaluated the health data processed in different standards, represented by *open*EHR and HL7/FHIR. OmniPHR demonstrated the feasibility to provide semantic interoperability through a standard ontology and machine learning with NLP (Natural Language Processing). Although 12% of health records still required manual intervention in conversion, we present a way to obtain the original data from different standards on a single format. We evaluated our model implementation using the data set of more than 40,000 adult patients anonymized from two hospital databases. We tested the distribution and reintegration of the data to compose a single view of health records. Moreover, we profiled the model by evaluating a scenario with ten superpeers and thousands of concurrent sessions transacting operations on health records simultaneously, resulting in an average response time below 500 ms. The Blockchain implemented in our prototype achieved 98% availability. **CONCLUSION:** As contribution, OmniPHR presents a unified, semantic, and up-to-date vision of PHR for patients and health providers. The results were promising and demonstrated the possibility of subsidizing the creation of inferences rules about possible patients' health problems and preventing future problems.

Keywords: Health Record. Blockchain. Semantic Interoperability. Natural Language Processing. Distributed Systems. Health Informatics.

RESUMO

CONTEXTO: Os avanços na Tecnologia da Informação trouxeram muitos benefícios para a área da saúde, especialmente para o armazenamento digital dos registros de saúde dos pacientes. No entanto, ainda é um desafio ter um ponto de vista unificado do histórico de saúde dos pacientes, porque normalmente os dados de saúde estão espalhados por diferentes organizações de saúde. Além disso, existem vários padrões para esses registros, alguns deles abertos e outros proprietários. Normalmente, os registros de saúde são armazenados em bancos de dados dentro das organizações e raramente se têm acesso externo. Essa situação se aplica principalmente aos casos em que os dados dos pacientes são mantidos pelas organizações de saúde, conhecidos como EHR (Electronic Health Record). No caso do PHR (Personal Health Record), no qual os pacientes podem gerenciar seus registros de saúde, eles geralmente não têm controle sobre seus dados armazenados nos bancos de dados das organizações. Mesmo com padrões de dados de saúde adotados, os pacientes muitas vezes precisam explicar diversas vezes suas informações de saúde quando são atendidos em locais diferentes. Esse problema dificulta a adoção do PHR. **OBJETIVO:** Desse modo, vislumbramos dois desafios principais no contexto de PHR: primeiro, como os pacientes podem ter uma visão unificada de seus registros de saúde dispersos e, segundo, como os profissionais de saúde podem acessar dados atualizados sobre seus pacientes, mesmo que as mudanças ocorram em outros lugares. A contribuição científica consiste em propor um modelo de arquitetura baseado em Blockchain para suportar um PHR distribuído, onde os pacientes possam manter seu histórico de saúde unificado, a partir de qualquer dispositivo e em qualquer lugar. Da mesma forma, a contribuição científica para os profissionais de saúde busca promover a possibilidade de interconexão dos dados dos pacientes entre as organizações de saúde. **METODOLOGIA:** A metodologia consiste em propor e prototipar um modelo de aplicativo chamado OmniPHR (Omni de onipresente) como um modelo distribuído para integrar os PHRs. Para avaliar o modelo, o método inclui avaliar desempenho da rede, interoperabilidade e integração semântica de diferentes padrões de saúde, usando um banco de dados real de pacientes anonimizados. **RESULTADOS:** As avaliações demonstram a viabilidade do modelo na manutenção de registros de saúde distribuídos em um modelo de arquitetura que promove uma visão unificada do PHR com escalabilidade da solução. Como resultado, avaliamos os dados de saúde processados em diferentes padrões, representados por *open*EHR e HL7/FHIR. O OmniPHR demonstrou a viabilidade de fornecer interoperabilidade semântica através de uma ontologia padrão e PLN (Processamento de Linguagem Natural). Embora 12% dos registros de saúde ainda precisem de intervenção manual na conversão, apresentamos uma maneira de obter os dados originais de diferentes padrões em um único formato. Avaliamos a implementação do nosso modelo usando o conjunto de dados de mais de 40.000 pacientes adultos anonimizados de dois bancos de dados de hospitais. Testamos a distribuição e reintegração dos dados para compor uma única visão dos registros de saúde. Além disso, analisamos o modelo avaliando um cenário com 10 super nós e milhares de sessões concorrentes transacionando operações em registros de saúde simultaneamente, resultando em um tempo médio de resposta abaixo de 500 ms. O Blockchain implementado em nosso protótipo atingiu a disponibilidade de 98%. **CONCLUSÃO:** Como contribuição, o OmniPHR apresenta uma visão unificada, semântica e atualizada de PHR para pacientes e profissionais de saúde. Os resultados foram promissores e demonstraram a possibilidade de subsidiar a criação de inferências sobre possíveis problemas de saúde do paciente e a prevenção de problemas futuros.

Palavras-chave: Registro de Saúde. Blockchain. Interoperabilidade Semântica. Processamento de Linguagem Natural. Sistemas Distribuídos. Informática em Saúde.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

AG	Architecture Group
AR	Archetypes
BR	Backbone Routers
BF	Blockchain Framework
CC	Challenge and Concern
DC	Distributed Components
DC	Dublin Core metadata
DE	Distributed Event-based
DO	Distributed Objects
GCC	Group of Challenges and Concerns
GQ	General Question
GRQ	General Research Question
GS	Group of Standards
HDS	Health Data Standard
MP	Messages Present
OHC	One-way Hop Count
OL	One-way Latency
ON	Ontologies
OmniPHR	Omnipresent Personal Health Record
RO	Routing Overlays
SA	Software Agents
SQ	Specific Question
SRQ	Specific Research Question
TM	Templates
TR	Terminologies

LIST OF ACRONYMS

ACM	Association for Computing Machinery
ACP	American College of Physicians
AI	Artificial Intelligence
API	Application Programming Interface
ASC	Accredited Standards Committee
CAN	Content Addressable Network
CCD	Continuity of Care Document
CCR	Continuity of Care Record
CDA	Clinical Document Architecture
CEN	European Committee for Standardization
CIM	Clinical Information Model
CNL	Controlled Natural Language
CS	Client-server
CSV	Comma-separated values
DHT	Distributed Hash Table
DICOM	Digital Imaging and Communications in Medicine
EHR	Electronic Health Record
EMR	Electronic Medical Record
ESB	Enterprise Service Bus
FHIR	Fast Healthcare Interoperability Resources
GATE	General Architecture for Text Engineering
HIMSS	Healthcare Information and Management Systems Society
HIPAA	Health Insurance Portability and Accountability Act
HIS	Hospital Information System
HIT	Health Information Technology
HL7	Health Level Seven
HNA	Home Nursing Activities
IADIS	International Association for Development of the Information Society
ICD	International Classification of Diseases
ICPC	International Classification of Primary Care
ICT	Information and Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IET	Institution of Engineering and Technology

ISO	International Organization for Standardization
IoHT	Internet of Health Things
IoT	Internet of Things
JBHI	IEEE Journal of Biomedical and Health Informatics
JBI	Journal of Biomedical Informatics
JMIR	Journal of Medical Internet Research
JSON	JavaScript Object Notation
KB	KnowledgeBase
LOINC	Logical Observation Identifiers Names and Codes
MIMIC	Medical Information Mart for Intensive Care
MLA	Medical Library Association
MTBF	Mean Time Between Failures
MTTR	Mean Time To Repair
NCBI	National Center for Biotechnology Information
NIC	Nursing Interventions Classification
NLP	Natural Language Processing
OWL	Ontology Web Language
P2P	Peer-to-Peer
PHA	Patient Health Application
PHD	Personal Health Data
PHI	Private Health Information
PHM	Private Health Management
PHR	Personal Health Record
PICOC	Population, Intervention, Comparison, Outcome and Context
PMC	PubMed Central
RDB	Relational Database
RDF	Resource Description Framework
REST	Representational State Transfer
SDK	Software Development Kit
SLR	Systematic Literature Review
SNOMED	Systematized Nomenclature of Medicine
SPARQL	SPARQL Protocol and RDF Query Language
SPASQL	SPARQL Protocol and SQL
SOAP	Simple Object Access Protocol

SQL	Structured Query Language
SWRL	Semantic Web Rule Language
UHR	Universal Health Record
XML	eXtended Markup Language
BioMed	BioMed Central
eHealth	Electronic Health
ePHR	Electronic Personal Health Record
iPHR	intelligent Personal Health Record
mHealth	Mobile Health
uHealth	Ubiquitous Health

CONTENTS

1 INTRODUCTION

Information and Communication Technology (ICT) has transformed the health care field worldwide. One of the main drivers of this change is the Electronic Health Record (EHR) (RAJKOMAR et al., 2018). However, there are still open issues and challenges because the EHR usually reflects the partial view of a health care provider. Besides, by definition, EHR does not address patients' ability to control or interact with their data (ISO, 2005). Furthermore, with the growth of mobile and ubiquitous computing, the number of records regarding personal health is increasing exponentially. This movement we characterize as the Internet of Health Things (IoHT), including the widespread development of wearable computing technology and assorted types of health-related sensors (COSTA et al., 2018). This area leads to the need for an integrated method of storing health-related data, defined as the Personal Health Record (PHR) (ISO, 2012), which could be used by health care providers and patients (ROEHRS et al., 2017). This approach could combine EHRs with data gathered from sensors or other wearable computing devices (ROEHRS; COSTA; ROSA RIGHI, 2017; ROEHRS et al., 2018). This unified view of patients' health could be shared with providers, which may not only use previous health-related records but also expand them with data resulting from their interactions (ROEHRS et al., 2019). Another PHR advantage is that patients can interact with their health data since their data are under their control (ISO, 2012).

The Health Information Technology (HIT) has evolved greatly, but even now, we generally have not our entire patient health history in a unified view (SARIPALLE; RUNYAN; RUSSELL, 2019). We still have different health records with assorted health care providers (i.e., health care professionals and health care organizations) that we interacted lifelong (VAN GORP; COMUZZI, 2014) (BOURGEOIS; NIGRIN; HARPER, 2015). At every medical appointment, patients must tell their whole health history again, losing time and accuracy. In addition, there are technical issues with health records, since there are several health data standards for different purposes, as can be seen in Table 1. The standards aim to systematize the patients' clinical datasets and define protocols to standardize health information. These are usually dedicated to standardize the storage and to regulate the clinical and demographic data about patients. Health records typically incorporate data regarding vital signs, laboratory exams results, evolution, and diagnosis. However, in some cases, the standards are guidelines designed to address health records in some regions or countries, such as standards CEN (LOZANO-RUBÍ et al., 2016) in Europe or xDT in Germany (MILSTEIN; BLANKART, 2016). Patient's health data are collected throughout life and can receive data from several sources, including health professionals records from laboratories, clinics or hospitals, including data from sensors that monitor the patient's health (HEINTZMAN; KLEINBERG, 2016) (MIHAJLOVIĆ et al., 2015).

The area of health information technology has evolved in the application of standards for health record definition, through the adoption of EHR (SHORTLIFFE; CIMINO, 2013). The purpose of EHR is to standardize health data, but without determining or specifying which

Table 1 – Standards for health records storage and communication.

Acronym (Reference)	Short Description
ASC X12N (ASC, 2017)	Accredited Standards Committee X12N
CCR (CCR, 2017)	Continuity of Care Record
CEN/TC 251 (CEN, 2017)	European Committee for Standardization
DICOM (DICOM, 2017)	Digital Imaging and Communic. in Medicine
HL7/CDA/FHIR (DOLIN et al., 2015)	Health Level-7 / Fast Health. Interop. Res.
HIPAA (HIPAA, 2017)	Health Insur. Portab. and Account. Act
ICD/ICF/ICHI (ICD, 2017)	Family of International Classifications
ICPC (ICPC, 2017)	International Classification of Primary Care
IHE (IHE, 2017)	Integrating the Healthcare Enterprise
ISO/TC 215 (ISO, 2017a)	International Organization for Standard
LOINC (LOINC, 2017)	Logical Observ. Identif. Names and Codes
*open*EHR (OPENEHR, 2017)	Open Electronic Health Records
SNOMED-CT (SNOMED, 2017)	Systematized Nomenclature Of Medicine
xDT (MILSTEIN; BLANKART, 2016)	Germany Family of Data Exchange Formats

Source: Prepared by the author

standard to adopt. Another way to obtain patients' health data in an electronic and equalized format is through the PHR (TANG et al., 2006; ARCHER et al., 2011). The ISO TR14639-2:2014 indicates that PHR is the "representation of information regarding or relevant to the health, including wellness, development, and welfare, of a subject of care, which may be stand-alone or integrating health information from multiple sources" (ISO, 2014).

The adoption of the EHR has evolved as a consolidated technology for recording patient health data (JAMOOM; YANG; HING, 2016; YADAV et al., 2018). A key difference between EHR and PHR is that PHR enables patients to access and control their data (ISO, 2017b). PHR is an emerging trend with growth potential in the health care domain (WASS; VIMARLUND, 2018). Improving the management and sharing of health records is a key focus of our work reported in this study. Although initiatives to adopt PHR have evolved in recent years, they face barriers to adoption (NEW et al., 2018). One barrier faced by both EHR and PHR is the distribution and limitations of health record integration. Other barriers relate to security issues, such as confidentiality and privacy of health records (FORD; HESSE; HUERTA, 2016; ALYAMI; SONG, 2016).

In this sense, Blockchain technology has come up with a strong appeal to the financial area, especially in the use of virtual currencies, also known as crypto-coins (NAKAMOTO, 2008). However, after a few years, potential uses of this technology began to emerge in other business areas (MCGHIN et al., 2019). Due to the premise of implementation based on data distribution in Peer-to-Peer (P2P) networks, Blockchain has opened up possibilities for use in other fields, such as health area (METTLER, 2016). Regarding this area of business, the particular highlight is the possibility of use for the integration of patients' health records (MCGHIN et al., 2019). In other words, as chained blocks can store records of transactions with electronic money, it was seen that the health records of patients could also be chained (AZARIA et al., 2016).

1.1 Motivation

The physician-patient relationship traditionally consists of the patient's dependence on the physician's recommendations (OROM et al., 2018). Physicians need to keep accurate record systems to store information about patients and use the records to make diagnoses and recommendations (RATHERT et al., 2019). In this sense, one significant milestone is the use of the EHR. Health records are collections of patient health data, and the EHR is defined as a digital repository of the health status of patients (SCHINASI et al., 2018). The EHR evolved from several electronic methods of storing patients' health data that became a structured and interoperable approach (CASTILLO; MARTÍNEZ-GARCÍA; PULIDO, 2010; ISO, 2005).

However, EHRs have some limitations because their records are based entirely on data reported by health care providers (BRENNAN; DOWNS; CASPER, 2010). One trend is allowing patients to have access to their health data, making them the owner of such data (MEIER; FITZGERALD; SMITH, 2013). Therefore, PHR arose from the EHR and is defined as a health record related to patient care, which is controlled by the patient. (SPIL; KLEIN, 2015; TANG et al., 2006). The PHR can also be defined as a representation of the health information, wellness, and development of a person (ISO, 2012). The main advantages of the PHR refer to the ability of patients to maintain data on their health. However, many challenges need to be overcome to promote widespread PHR adoption, including how to achieve interoperability using the EHR, implementation costs, privacy, security, and the assessment of the effective benefits that the patient may have (BAIRD; NORTH; RAGHU, 2011).

The PHR works as a platform for patients' and health care providers' use, enabling the exchange of information with health care systems (YAO et al., 2018). PHR has also emerged as a mechanism for patients to make appointments with their health care providers. The aim is to address patients' evolving needs by using specific methods to improve their care and foresee health issues. The technologies used to process health-related data include machine learning, pattern recognition, applied mathematics, statistics, expert systems, data sharing, and artificial intelligence algorithms (ANDREU-PEREZ et al., 2015). Moreover, advances in ICT have allowed both the storage and easy access of large amounts of data, allowing the release of physical space, facilitating research and the correlation of data within hospitals (YAO et al., 2018). However, the increasing number of patients who need care, especially with the increased life expectancy of people in several countries, has been an obstacle to managing huge databases of medical records. The health community is constantly facing global epidemics and issues that transcend countries, such as cancer, influenza, AIDS, diabetes, and obesity. Patients who migrate or travel from one country to another could make use of their PHR to obtain faster and more efficient health services. With the increase in the adoption of wireless technology and mobile devices, this creates opportunities to deliver health care services to patients through a world-standard PHR, although many challenges remain in achieving these benefits (WELLS et al., 2014).

In this way, Blockchain technologies (NAKAMOTO, 2008; ZHANG et al., 2017) are a promising means to address the barriers with distributed PHRs described above by forming a unified view of PHRs. Blockchain technology has been researched and implemented in various domains, initially in the financial domain with virtual currencies and more recently in the health domain (RANDALL; GOEL; ABUJAMRA, 2017; DAGHER et al., 2018). Various approaches to applying Blockchain to health data have been proposed, centered mostly around composing a distributed ledger of health records (KUO; KIM; OHNO-MACHADO, 2017) and providing useful tools to preserve patient privacy (GOLDIM; GIBBON, 2015). The performance of distributed PHRs and integration of health data among health organizations are crucial factors for ensuring the adoption of Blockchain technologies.

The main feature of the distribution of patient records, forming a complete and unique history, fits in Blockchain model since many health providers attending to a patient over a lifetime (CHEN et al., 2019). A data-chained distribution model facilitates the application of this technology to the model of patient health records and can form a single view of these data. Another factor that fits the Blockchain technology model is the fact that health records do not follow a centralized model, and in this sense, health records can actually belong to the patient, shaping the concept of PHR (PUTHAL et al., 2018). Another aspect of the use of Blockchain, especially regarding health records, is that these data are the fundamental basis for making it possible to expand medical use. In this case, Blockchain intends to form the patient's complete history, allowing for the distributed processing of any manipulation or query of the data. In addition, Blockchain technology allows adding security features, given its resilient nature of data validation across all network nodes and the encryption capabilities to meet confidentiality assumptions (CACHIN; VUKOLIĆ, 2017).

PHRs allow patients to maintain information on their medical conditions, drugs, and behaviors related to self-care and self-monitoring of their health (CUNNINGHAM; AINSWORTH, 2018). Nevertheless, access controlled by the patients represents an ever-present concern because it requires a free but safe balance between system customization, privacy, and security controls (LIU; HUANG; LIU, 2015). In particular, without the application of security practices, no privacy is available for the data (OZOK et al., 2014). Another possibility is that the PHRs accept data obtained from health-related equipment, such as accelerometers, gyroscopes, wireless scales, wristbands, and smartwatches. The proliferation of these technologies is called the Internet of Things (IoT) (LI; DA XU; ZHAO, 2015). Among IoT application domains, health care is one of the most attractive, giving rise to many health-related devices (ISLAM et al., 2015). Data collected from these objects can complement the PHRs and help detect risks to the patients' health (GUBBI et al., 2013). Nonetheless, existing PHRs have limited intelligence and can only inform a small subset of users' health care needs (LUO; TANG; THOMAS, 2012). Besides, processing PHR data automatically and combining data from sensors with stored records for transformation into useful knowledge is another challenge (BLAKE, 2015). In conjunction with Blockchain technology, and to compose the patient's medical record, we have the Internet of Health Things (IoHT) concept (COSTA et al., 2018). IoHT aims to aggregate to the patient's record the collection of data in real-time of the patient's health, such as data of online monitor of the health status. In this way, Health Information Systems (HIS) can collect patients' health data in several places and HIS can feed the patient's record constantly with up-to-date data. Therefore, with the formation of the complete and constant history of the patient's records, new possibilities open to analyze these data, helping in the formation of the Big Data of the patient's health records (WANG; KUNG; BYRD, 2018).

1.2 Problem

There are several health data standards (FRAGIDIS; CHATZOGLOU, 2018). Many health care providers adopt proprietary standards without integration with others. In some countries, there are recommendations for adopting recognized health data standards. One of the main goals of using standards is to provide interoperability among health care organizations. Nevertheless, using open and internationally recognized standards does not guarantee interoperability because many of them are incompatible with each other (MANDEL et al., 2016). In this sense, the patients' data are difficult to integrate (EDEN et al., 2016; CANTOR; THORPE, 2018). Even with the evolution of open specifications and attempts to promote the use of the standards, the adoption of EHR/PHR is still challenging (FORD; HESSE; HUERTA, 2016). In the PHR case, which can also aggregate data from wearable devices of the patient, the integration can be more complex. This is because the PHR aims to gather all the patient's health data, regardless of the health care provider (FRICTON; DAVIES, 2008). Besides, the syntactic standards have limited benefits because their overall purpose is only to structure or standardize the format and terminologies used in the health records (VUOKKO et al., 2017). In summary, the problem statement of this work is regarding the difficulty to integrate the several existing standards of patient health data.

Many health systems use databases in proprietary formats. These databases are structured to be accessed exclusively by those systems, with little or no interoperability with others (KRAAN et al., 2015). Usually, legacy systems in many health organizations preserve proprietary data structures. In general, these databases are hosted in a data center inside the health organizations, with restricted access to internal health professionals. In some cases, e.g., laboratory exams results, patients and health care providers can have external access to health records in a restricted manner, only to be viewed or printed. Another factor is that health data is becoming increasingly larger. Several studies bring out crucial points as getting this mass data about patients' health, such as standardization of data, storage capacity, location, safety, and how to filter, analyze and quickly obtain such data (O'DRISCOLL; DAUGELAITE; SLEATOR, 2013). Allied to these issues, health organizations maintain the patient's EHR indefinitely, even outdated. This is required for legal reasons, depending on the country (BOURGEOIS; NIGRIN; HARPER, 2015). PHRs have a problem of health data distribution since, in many cases, health care providers do not share their patients' data. Hence, they do not have these data up-to-date when other health care providers assist their patients (DYE et al., 2016). Moreover, these records are usually stored in different standards on different health organizations, which bring difficulties for exchange health records between organizations (BHARTIYA; MEHROTRA; GIRDHAR, 2016). Besides, there are several health standards for different purposes and initiatives to mitigate some integration problems among health systems (MARCOS et al., 2015).

Other problems arise from the potential existence of health records duplicated within the health organizations due to the ambiguity or repetition of some patient's names (MCCOY et al., 2013) (KRAAN et al., 2015). Furthermore, from the patients' viewpoint, they do not have an integrated view of their health records. Although there are consolidated standards to structure the patient's health data, the adoption and implementation of EHR, particularly PHR, is still a challenge (SIMPSON, 2015). Much of the obstacles come from the fact that health records are sensitive and have complex management for owners and users (LI et al., 2013) (ISTEPHAN; SIADAT, 2016). There are concerns in PHR adoption from health care providers and patients because users are afraid to share their data, as there are concerns about where data will be stored and who will have access to it (TONG et al., 2014). Other barriers include concerns from health care providers regarding the management and validity of records registered in PHR since patients are the owner and can manage their records (KRAAN et al., 2015). In addition, because of the high cost of datacenters, many PHR services have migrated to third party providers using cloud computing architectures (LI et al., 2013). However, according to Mxoli (MXOLI; GERBER; MOSTERT-PHIPPS, 2014) "access management, security issues, legal issues and loss of data are some of the risks that negatively affect the storing of PHRs in the Cloud" (MXOLI; GERBER; MOSTERT-PHIPPS, 2014).

Patient health data are conventionally stored in health care provider repositories (HEART; BEN-ASSULI; SHABTAI, 2017; GARDIYAWASAM PUSSEWALAGE; OLESHCHUK, 2017). Often, however, these data are not shared between providers or with patients. Moreover, even where there is an intention to share data, there are barriers to achieving this goal (SHOWELL, 2017), including

(a) Interoperability stemming from the lack of common health data standards (ALYAMI; SONG, 2016).

(b) The difficulty of integrating large amounts of data contained in medical records (KAUR; RANI, 2015).

As a consequence, patients must often re-inform their health history, repeat laboratory exams, or even perform unnecessary tests when they are attended by different health providers (KRASOWSKI et al., 2015). Although some countries have initiatives to integrate personal health history, this integration often occurs only at the organizational level, without patients having access to their digital records (FRAGIDIS; CHATZOGLOU, 2017). In such cases, therefore, only the data reported in the health organizations are integrated, regardless of factors like patient wellness data, nutrition, data collected on wearables, or collected on monitoring equipment at home (CHIAUZZI; RODARTE; DASMAHAPATRA, 2015). Moreover, patient care often comes from health providers who are not part of an integrated network of health organizations, e.g., if patients are treated in a foreign country (GARDIYAWASAM PUSSEWALAGE; OLESHCHUK, 2017).

Since its inception in the past decade, Blockchain's technology-driven field of patient health records has shown potential (PARK et al., 2019). Since then, various initiatives, proposals, and use models have emerged. At the same time, various tools are emerging to facilitate the implementation of Blockchain technology for general purposes. However, even with this evolution, in fact, still a few solutions have implemented Blockchain applied to health records in production environments (PARK et al., 2019).

1.3 Research Question

We researched the recent scientific literature to identify issues and barriers regarding PHR. The objective was to identify gaps and opportunities about PHR, as well to determine the main research question that underlies our study. Considering the issues, barriers and challenges facing the adoption of PHR previously presented, specifically regarding problems with interoperability and health data distribution, the research goal underlying this study is to answer the following main research question:

> *How would it be possible to have a single view of a PHR being distributed, up-to-date, and semantic interoperable to patients and health care providers use?*

As explained throughout this study, our focus is on addressing problems related to the distribution and semantic interoperability of patient health data. In this sense, we consider as distribution problems the fact that the patients' health data are spread in several health organizations. We consider semantic interoperability problems the fact that several standards of health data are incompatible with each other, which prevents the extraction of knowledge from them. Moreover, we consider as a problem of updating the health data, the consequent fact that, besides to the problems cited above, patients do not have a unified and up-to-date view of their data, as well many health organizations do not share their patient data.

1.4 Scientific Contributions

To identify the technology for the PHR and to discuss the main open issues, this work started surveying the main contributions of the scientific community over the last decade. The purpose was to review the PHR literature and describe the existing models. As a way of mapping this scenario, we used the Systematic Literature Review (SLR) methodology to choose the studies (KITCHENHAM; CHARTERS, 2007), (KITCHENHAM; BRERETON, 2013), (GARCÍA-

BORGOÑON et al., 2014). As a contribution of this first part of the study, we propose an updated and wide taxonomy for PHRs and indicate further directions for study (ROEHRS et al., 2017).

The main scientific contribution is to provide a distributed and interoperable architecture model using the Blockchain technology for PHR, which addresses a unified viewpoint for both patients and health care providers. Patients can take advantage of maintaining their health history in a single view, as well as health care providers have these data up-to-date, regardless of where the patient was treated. To answer the research question, we propose a model named OmniPHR, where the prefix 'Omni' comes from omnipresent, meaning that is present everywhere (ROEHRS; COSTA; ROSA RIGHI, 2017).

A key aspect of our work involves evaluating a model for distributed PHR integration based on Blockchain technology (NAKAMOTO, 2008; TAPSCOTT; TAPSCOTT, 2016). The research gap that our work addresses involves determining how to develop a distributed and interoperable PHR implementation using Blockchain technology to integrate patient health records (ROEHRS et al., 2019). In particular, this work:

(a) evaluates the distribution and reintegration of health records via Blockchain technologies to compose a unified PHR view,

(b) analyzes the assessment of non-functional performance requirements, such as measure response time, CPU usage, memory occupation, disk and network usage of a varied number of superpeers and concurrent sessions transacting different operations on health records simultaneously, and

(c) discusses best practices for deploying Blockchain technologies in health care.

The OmniPHR approach is innovative since it promotes the integration of health data through the use of a distributed, private, and customizable platform, along with interoperable and standards-based protocols (ROEHRS; COSTA; ROSA RIGHI, 2017; ROEHRS et al., 2018, 2019). Likewise, we integrate distributed health records in a unified, safe, and interoperable manner for use by health providers and patients. In particular, the essential contribution is that OmniPHR promotes the sharing of PHRs among health care providers, with the possibility of knowledge and consent of the patient.

Therefore, this study aims to promote the formation of an integrated PHR as a basis for full clinical knowledge with the use of Blockchain technology in a differentiated setup from the traditional one. In this way, this study aims to address as contribution two main aspects:

(a) The first contribution aims to propose a disruptive business model architecture based on Blockchain technology, to promote the implementation of a complete and distributed health record of the patient.

(b) As a second contribution, this proposal aims to shed light on the integration of different Blockchain-based architectures, i.e., specifically regarding the orchestration of multiple Blockchains (Multi-Blockchain), in order to make the integration of patients' health records widely possible.

As contribution regarding the interoperability, the proposal presents an application model to addresses the integration issues between health data standards, providing a single, semantic and up-to-date PHR viewpoint, through ontology and Artificial Intelligence (AI) using Natural Language Processing (NLP) to automate the conversion of different health standards (ROEHRS et al., 2018). According to the problem statement of health data interoperability, the objective is to propose an architecture model that enables the improvement of health services for patients and health care providers through the semantic interoperability of health data. These objectives aim to enable continuous updates on the PHR, independently of the places where the patient has their data collected or changed.

1.5 Study Organization

The remainder of the study is divided as follows:

(i) Chapter 2 summarizes the main concepts, challenges, and models that support the proposal, detailing the concepts of EHR, PHR, as well as the respective advantages and disadvantages.

(ii) Chapter 3 explains the methodology applied in each stage of the work, including the methods used in the systematic literature review, architecture proposal, and semantic interoperability proposal.

(iii) Chapter 4 presents the main related work and the strategies to select them.

(iv) Chapter 5 presents the foundation technologies for model development and details the architecture of OmniPHR model.

(v) Chapter 6 presents the implementation details, evaluation applied in each stage, and summarizes the results obtained.

(vi) Chapter 7 discusses the impacts, opportunities, research limitations, and future directions.

(vii) Chapter 8 presents the final considerations regarding the findings and future work.

2 BACKGROUND

This chapter summarizes the terminology and platforms used in this work. This chapter presents the main concepts in order to support the proposed solution, as well as important definitions for the understanding, classification, and organization of the proposed work. We start with the concepts, advantages, and disadvantages of EHR, PHR, and the definition of interoperability. Afterward, we describe the technologies that complete the solution and how they are interconnected with the model, including the technologies: Blockchain, Routing Overlay, *open*EHR health data standard, and Chord algorithm.

2.1 Electronic Health Record - EHR

The EHR, also called electronic medical record, refers to a structure in digital format of patients' health data that is maintained throughout their life and is stored accurately in a repository (GUNTER; TERRY, 2005). Health care providers use EHRs, whose data can vary greatly and can include vital signs (such as body temperature, pulse, respiration, and blood pressure), age, weight, medications, allergies, medical examination results, and radiology images that are used to diagnose conditions (GUNTER; TERRY, 2005), (ISO, 2005). The EHR is used to support health care professionals and health organizations (e.g., hospitals, laboratories, or clinics) for the improved management of patient health data (CALIGTAN; DYKES, 2011). However, these health records are usually not stored with the same structure in different health organizations. These factors hinder the interoperability of health information among hospitals, clinics, and laboratories (ALABBASI et al., 2014). To address some of these problems, the PHR concept was proposed in 2006 (TANG et al., 2006) and was defined as an ISO (International Organization for Standardization) standard (ISO/TR 14292) in 2012 (ISO, 2012).

According to ISO/TR 14639, EHR is “information relevant to the wellness, health and health care of an individual, in computer-processable form and represented according to a standardized information model" (ISO, 2014). EHR refers to a structure in an electronic way of patient's health records, collected and stored in a repository, which can be shared by different digital formats. EHR can contain several data groups, such as allergies, vital signs, medical appointments, laboratory exams results, medical imaging, and diagnoses. To differentiate health records that are not integrated between health care providers, these are named EMRs (Electronic Medical Records). EMR can be considered a special type of EHR with specific focus into the internal medical domain of health organizations (ISO, 2014) (HEART; BEN-ASSULI; SHABTAI, 2016).

2.1.1 Advantages and Disadvantages of EHR

EHR is a standardized information model, enabling integration among multiple health care providers, and this integration is considered their main advantage (ISO, 2014) (HEART; BEN-ASSULI; SHABTAI, 2016). EHR has several benefits, ranging from supporting medical prescriptions (CHEN, 2016), improving disease management (ROUMIA; STEINHUBL, 2014) and contributing in the reduction of severe medication errors (HAN et al., 2016).

However, EHR has limitations regarding interoperability, e.g when health organizations adopt international but heterogeneous standards (BHARTIYA; MEHROTRA; GIRDHAR, 2016). Other limitations are related to the security of data exchanged between health organizations, or to non-incorporation of data about patient's wellness, such as sports activities or eating habits (CHEN, 2016). According to the ISO definition (ISO, 2014), EHR aims at standardizing health data, but without determining or specifying which standards should be adopted. In this sense, it is up to the health providers to choose which standards to use in their health organizations.

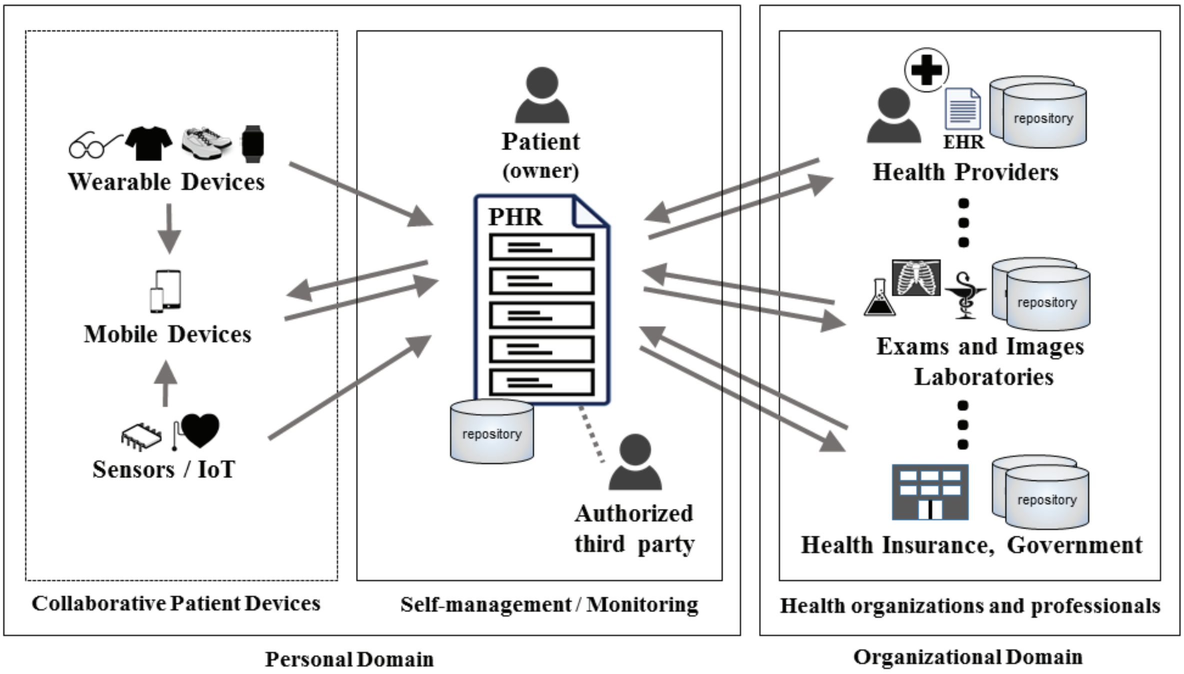
2.2 Personal Health Record - PHR

As an evolution of EHR, we have the concept of PHR. PHR refers to a representation of health records related to the care of a patient that is managed by the patient (TANG et al., 2006). In other words, the PHR refers to archives containing health data about each patient, but, unlike the EHR, it is managed by the patient (BAIRD; NORTH; RAGHU, 2011), (ISO, 2012). With a PHR, patients can choose to share their health data with health care providers or keep them private (TANG et al., 2006). Figure 1 illustrates how the PHR and EHR differ in their goals, although they can be integrated to exchange information that is relevant to the patient's health (ISO, 2012).

According to ISO/TR 14639, PHR refers to a "representation of information regarding, or relevant to, the health, including wellness, development, and welfare of that individual" (ISO, 2014). As patients are the owner of their health records, they can manage and grant permissions for access or share their health data with third-parties (ISO, 2014). PHR is oriented to the patient but can be integrated with EHR (ROEHRS et al., 2017). Some health care providers have been successful in improving communication with patients using mobile technology (mPHR), where PHR allows patients self-monitoring and managing their health status (REEDER; DAVID, 2016). PHR can receive data from health care providers, stored in a repository where the patient has access (ARCHER et al., 2011).

Some variant names for PHR appeared in the literature, such as ePHR (electronic PHR) (WYNIA; DUNN, 2010) or UHR (universal health record) (MORGENTHALER, 2011). The first concept refers to the use of PHR in an electronic format, while the second proposes PHR-sharing data with health care providers. Another term is intelligent PHR (iPHR), which uses medical knowledge to anticipate the health needs of patients and promote tools to guide searches

Figure 1 – PHR and EHR relationships.



Source: Prepared by the author.

for diseases and recommendations for nursing activities or medical products (LUO; TANG; THOMAS, 2012). Although these different nomenclatures are used, we use the term PHR throughout this work.

2.2.1 Advantages and Disadvantages of PHR

Multiple EHRs for the same patient can coexist, but only one PHR would exist. The PHR can integrate data from many sources, ranging from devices connected to the patient to data from EHRs stored in health care provider systems (TANG et al., 2006). Although PHR may refer to records regardless of format (and can be on paper), the records are implemented electronically and are accessible through mobile devices (mHealth). Therefore, PHRs have allowed patients to self-monitor and manage their own health conditions (HORAN; BOTTS; BURKHARD, 2010). Another alternative is medical-oriented PHR, which includes features that are not patient-centered (FUJI et al., 2012), (TANG; LANSKY, 2005). This PHR can be "tethered" (tied) to where the data subsets are provided, including organizations that maintain patient data electronically (TANG et al., 2006). Hence, PHRs may be stored in a stand-alone computer or service portal to which only the user has access (ARCHER et al., 2011).

PHR has some advantages over EHR since PHR can receive data entered by the patient (ROEHRS et al., 2017). For instance, the patient can inform weight or blood pressure readings (GEORGE; HOPLA, 2015). However, PHR has some limitations and challenges (ROEHRS et al., 2017). PHR issues range from usability (as usefulness, satisfaction, and ease of use)

(WANG; DOLEZEL, 2016); low level of adoption (e.g. by patients with chronic medical conditions) (SHIMADA et al., 2014); few patients and physicians knowledge regarding PHR features; incompatibility or lack of integration with existing health systems; to concerns with security and access permissions for third-parties (e.g. physicians and family members) (BUTLER et al., 2013).

2.3 Interoperability

The concept of interoperability is quite broad and applied in many contexts (HOGAN et al., 2014). According to the Healthcare Information and Management Systems Society (HIMSS) definition (HIMSS, 2018), there are three levels of health data interoperability: (a) foundational, which makes the exchange of data between health systems possible without requiring the ability to interpret the data; (b) structural, which defines the syntax for data exchange, ensuring that data interoperability can be interpreted at the data field level; and (c) semantic, which "takes advantage of both the structuring of the data exchange and the codification of the data including vocabulary so that the receiving information technology systems can interpret the data" (HIMSS, 2018). Semantic interoperability ensures that systems understand data in the same way, resulting in unambiguous use, understanding, and interpretation of the data (SOCEANU, 2016). Semantic interoperability brings, besides the standardization and formatting of health data, the possibility of inferring based on the data. Instead, syntactic interoperability refers to the dealing of data with low-level problems, such as in the use of different protocols and formats (MARCOS et al., 2015).

Reviewing the literature, we identify some techniques that can help in the semantic interoperability of health data, with the individual or combined use of:

(a) Dublin Core metadata standard (DC) (ALYAMI; SONG, 2016) (SONG et al., 2017) - where the metadata could be used to describe, retrieve and organize the document with health records that do not follow the open standard;

(b) Natural Language Processing (NLP) (OEMIG; BLOBEL, 2014) (MALIK; SALEEM, 2016) - where NLP could be used to help the parser of legacy contents to a standard format, adding the possibility of extracting knowledge from the health records;

(c) Ontologies (MANDEL et al., 2016) (ESPOSITO; CASTIGLIONE; PALMIERI, 2016) - where representations through ontologies could be used to compose a standard that mediates heterogeneous standards, adding the possibility of extracting inferences from this composition;

(d) Software Agents (MORAES et al., 2016) (HU; ELKUS; KERSCHBERG, 2016) - where agent-based interface systems are designed to interpret the health records.

2.4 *open*EHR Standard

As the purpose of this proposal is to promote the integration of patient health records throughout life, we use international standards for the structuring of health data. These standards aim to promote the interoperability of health records and are key to designing a unified view of the patient's chart. In this sense, there are several standards for different parts of what can form a complete medical record, such as the set of norms for use of DICOM (Digital Imaging and Communications in Medicine) (GONÇALVES-FERREIRA et al., 2019), SNOMED-CT (SNOMED Clinical Terms) (TEODORO et al., 2018), and the LOINC (Logical Observation Identifiers Names and Codes) (WULFF et al., 2018), among others. In addition, more broadly, in standardizing the format and high-level structure of health records, there are several formats and protocols around the world. Two of the most recognized international standards used in various countries to structure interoperable health records are HL7 (SARIPALLE; RUNYAN; RUSSELL, 2019) and *open*EHR (YANG; HUANG; LI, 2019). These standards aim to provide the structural format of patients' records, as well as integrate with specific formats, such as DICOM (GONÇALVES-FERREIRA et al., 2019), SNOMED-CT (TEODORO et al., 2018), and LOINC (WULFF et al., 2018) standards.

The HL7 with FHIR and *open*EHR are among the main structural patterns of health data standards (BENSON; GRIEVE, 2016). Analyzing the main data standard regarding health records to be used in our proposal, *open*EHR (OPENEHR, 2017) stands for promoting a flexible structure based on archetypes. A key requirement for interoperability and important feature is that *open*EHR connects with others health data standards, such HL7 (DOLIN et al., 2015), LOINC (Logical Observation Identifiers Names and Codes) (BELLAMY, 2016), SNOMED-CT (Systematized Nomenclature of Medicine - Clinical Terms) (MORTENSEN et al., 2015) and DICOM (PANDIT; BOLAND, 2015). Moreover, the archetypes format of *open*EHR follows the premise of datablocks, which fits the OmniPHR purpose of having health datablocks chained on a P2P network.

One way to make health records interoperable is to use recognized data standards or protocols (MANDEL et al., 2016; SACHDEVA; BATRA; BHALLA, 2017). Several health data standards are defined around the world, with different purposes. Two internationally recognized standards used for electronic medical records are HL7 (ALIAKBARPOOR; COMAI; POZZI, 2017) and *open*EHR/ISO CEN13606 (ULRIKSEN; PEDERSEN; ELLINGSEN, 2017). The *open*EHR standard has the differential to treat health records semantically through ontology (LEGAZ-GARCÍA et al., 2015). In the *open*EHR standard, instances of datablocks can be serialized in either archetype (RDF/XML or JSON) or ontology (OWL) format, where RDF stands for "Resource Description Framework" and OWL stands for "Web Ontology Language."

2.5 Distributed Architectures Models

Considering that our subject of study is related to an architecture model for PHR based on a distributed system, we look for the main models mentioned in the literature. According to the classification of Coulouris (COULOURIS et al., 2011), there are five possibilities of architecture models:

(i) CS (Client-server), where “client processes interact with individual server processes in potentially separate host computers to access the shared resources that they manage";

(ii) P2P (Peer-to-Peer), where “all of the processes involved play similar roles, interacting cooperatively as peers without any distinction between client and server";

(iii) DO (Distributed Objects), where “each process contains a collection of objects, some of which can receive both local and remote invocations, whereas the other objects can receive only local invocations";

(iv) DC (Distributed Components), where “application servers provide structure to support a separation between application logic and data storage";

(v) DE (Distributed Event-based - services), where “the essence of indirect communication is to communicate through an intermediary and hence have no direct coupling between the sender and the one or more receivers."

And, still according to Couloris (COULOURIS et al., 2011), P2P systems are a trend for distributed systems because they have storage capacity and resource sharing on a global scale, but they have as a limitation the management and provision of adequate access to all the load to which they are subject. In addition, to compose the concept of P2P, Couloris (COULOURIS et al., 2011) defines the concept of routing overlay, which is described in the following.

2.6 Routing Overlay

In a P2P network, there is the concept of routing overlay, also known as superpeer or ultrapeer, which have special functions in a distributed system (COULOURIS et al., 2011). A routing overlay network aims to decentralize data and locate nodes on the network, managing their location. This mechanism has some certain goals, such as providing distribution, replication, security, and privacy. In our proposal, the health records are broken into small pieces distributed and encrypted on the network. The routing overlay must have special skills to manage responsibilities such as: (a) maintain system user registers; (b) keep PHR data, including new and update datablocks; (c) querying datablocks to assembly PHR when required; (d) maintain access permissions to health records; and (e) maintain access profiles to health records. In addition to these responsibilities, the routing overlay application needs to have functions granted to system administrators, such as the capabilities to maintain: (a) types of profiles; (b) health datablocks inherent in the standard; and (c) other interconnected standards of health records. In this sense, we have the Blockchain concept as a technology capable of unifying the concepts of distributed systems to support our proposal, which is described in the following section.

2.7 Blockchain Technology

Blockchain was first proposed to serve as a backbone for the Bitcoin technology, known as Satoshi's bitcoin model (TAPSCOTT; TAPSCOTT, 2016). In Bitcoins case, the coin is composed of distributed transactions chain in the network, which uses the principle of a P2P network to concentrate data in a single location (NAKAMOTO, 2008). In this sense, a block contains data regarding transactions and about the previous block, which links to the first block when the Bitcoin network started (NAKAMOTO, 2008). Blockchain is formed by a distributed database, which maintains a chain of datablocks, hence the origin. Each datablock refers to another within the block list, forming a complete chain, from first to the last datablock. These datablocks are distributed in a P2P network, making it difficult to manipulate this data by attackers. Applied to Bitcoin, the public key encryption mechanism is used to ensure the security of the electronic currency. This cryptography type is based on algorithms that require two keys, one public and the other private. The Bitcoin electronic currency is based on a chain of digital signatures with a central authority that verifies the validity of the chain (NAKAMOTO, 2008). In this case, the public key is used only to verify the digital signature applied to the transactions datablocks. More accurately, each datablock into the end of the chain is digitally signed and point to the next block using the public key of the latter.

Blockchain is a linked list of datablocks chained together in a distributed ledger by pointers, represented by a hash code that identifies each block, and where each datablock has, beyond the content, the pointer to the previous datablock in the chain (NAKAMOTO, 2008; NARAYANAN et al., 2016). In a Blockchain, each node in the peer-to-peer (P2P) network acts as a recorder of datablocks and as an evaluator of appropriate access and permissions of the content. Each node can add new blocks in the list and execute evaluation rules every interaction. These checks are performed in conjunction with the other nodes, forming the consensus protocol (KRAFT, 2016; STAGNARO, 2017).

Smart contracts are another concept applied in Blockchain technology to incorporate business rules or scripts to the processing performed on the platform. According to (SZABO, 1996), the smart contract is a "set of promises, specified in digital form, including protocols within which the parties perform on these promises." In many cases, smart contracts are used to verify the validity of contracts between two or more participants in a contract.

2.8 Chord Algorithm

In order to maintain the distribution of datablocks in an equitable manner throughout the network, OmniPHR can use an algorithm for P2P networks with Distributed Hash Table (DHT), such as CAN (Content Addressable Network), Chord, Kademlia, Pastry or Tapestry (COULOURIS et al., 2011). These algorithms can ensure equal distribution and knowledge of nodes where datablocks are located. To distribute health records parts, OmniPHR proposes the use of the Chord algorithm (STOICA et al., 2001), which is widely accepted on P2P networks (VATSAVAI; SURAVARAPU; MIR, 2016).

The goal is to get scalable (handling with increase amount of workload) and elastic (adapting to changes of workload) search service for OmniPHR distribution on the P2P network. The reasoning for selecting this algorithm is that Chord has an efficient node location and provides a balanced system in P2P networks (WOUNGANG et al., 2015). This suits the need to maintain the PHR datablocks chain using the principle of lookup in

$$O(\log N)$$

where N is the number of nodes in the network and grows logarithmically with the number of nodes. Chord operates two additional structures: finger table and successor. The algorithm uses a structure in which nodes are kept close, so a node is in the middle distance, another node a quarter off, and so on in a sequence powers of two. Chord algorithm promotes scalability, elasticity, availability, and durability of records. As can be seen in Figure 7, the purpose is that each part of OmniPHR may also be replicated in other nodes, distributed and chained on the network. Chord uses a variant of consistent hashing for load balancing, effecting in a uniform distribution (STOICA et al., 2001) (WOUNGANG et al., 2015).

Chord algorithm has flexibility and automatically adjusts the control tables according to enter and leave of nodes in the network (VATSAVAI; SURAVARAPU; MIR, 2016). Chord finger table is present at each node and has information about its identifier and IP address (WOUNGANG et al., 2015). The finger table contains data only on some near nodes, according to the execution of the algorithm, being the first table entry always refers to the successor node (STOICA et al., 2001). In the management of nodes, an important factor for OmniPHR regards to the replication of health datablocks. In the Chord algorithm, the list of successor nodes works as an engine that allows replicates data. When a regular node enters or leaves the network, a routing overlay is notified and knows whether it should disseminates copies. As a way to integrate and facilitate communication between nodes, we can use the precepts of the publish-subscribe system.

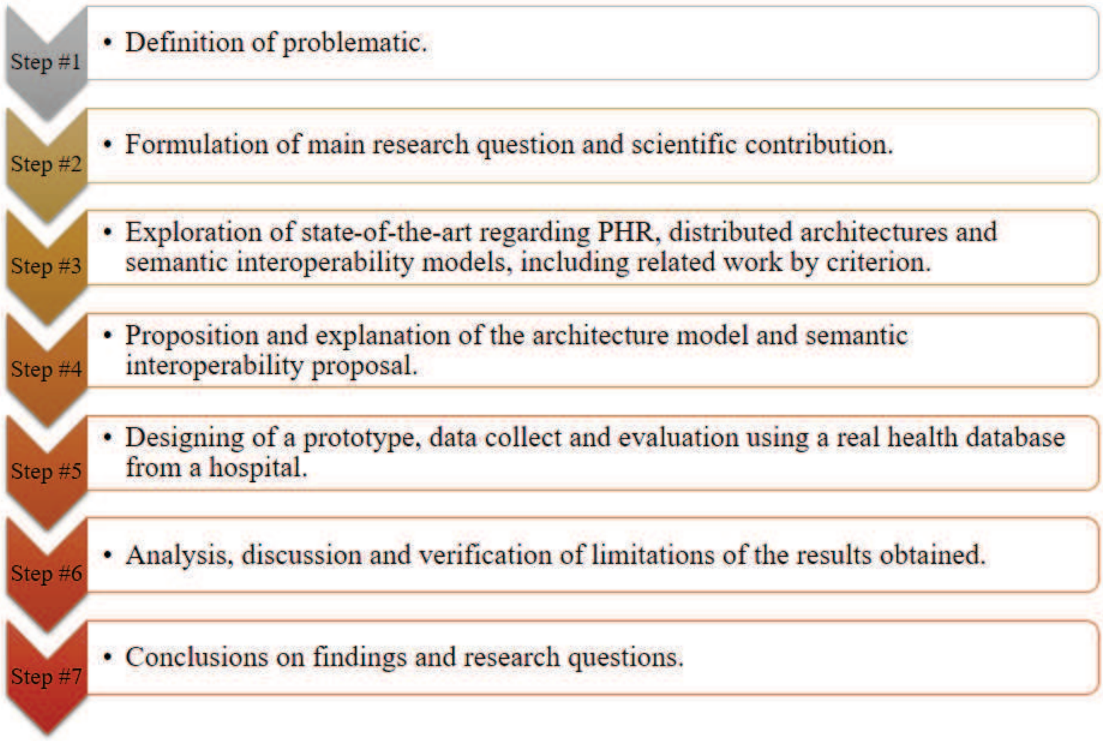
Nodes publishers publish messages to a service, represented in OmniPHR by routing overlay, and subscribers can get these messages, in an indirect communication between nodes (COULOURIS et al., 2011). Routing overlay application has the ability to receive and update datablocks containing information about PHR. Nodes publish messages with updated datablocks in the service addressed to certain nodes according to the DHT algorithm, and these nodes subscribe to their respective messages.

3 METHODOLOGY

In this chapter, we present step-by-step the scientific methodology used to support our research. The methodology proposed follows the principle used in the scientific community of designing a model, with the evaluation of a prototype system (GROENEN et al., 2016) (MAHER et al., 2016). The objective of this evaluation is to meet the health records distribution and interoperability requirements proposed for PHR.

Regarding the type of research, the approach is quantitative, since the analyzed health data is from existing patients, although with anonymized data. As for the nature of the research, the study is an applied research, since it aims at practical applications in the day-to-day of patients and health care providers. Regarding the objectives, the research was based on a case study applied to the context of the proposed model. Regarding the conduction, the research was divided into seven steps, described in Figure 2.

Figure 2 – Methodology.



Source: Prepared by the author.

According to the methodology presented, the first two steps have already been defined in the introduction of the study. The rest of the work is divided as follows.

(i) Section 3.1 explores how to achieve the state-of-the-art regarding PHR, and for this purpose, we followed the SLR methodology.

(ii) Section 3.2 presents specific aspects of the methodology to support and evaluate the architecture proposal.

(iii) Section 3.3 describes the methodology to evaluate the performance of the model.

(iv) Section 3.4 presents the methods to evaluate the Multi-Blockchain proposal.

(v) Section 3.5 explores the methods to evaluate the semantic interoperability proposal.

3.1 Exploration of the state-of-the-art regarding PHR

This section presents the methods used to identify the state-of-the-art in PHR through the SLR methodology. The details of the SLR methodology application and the results obtained are described in the next chapter of related work. We published a full article (ROEHRS et al., 2017) contemplating this methodology and the results obtained in the Journal of Medical Internet Research (JMIR).

3.1.1 Study Design

This section focuses on describing the study protocol, which introduces the adopted procedures and outlines the main subsequent decisions. As previously mentioned, this study presents the SLR method designed to provide a wide overview of the PHR research area, establish whether research evidence exists on a topic, and provide quantitative evidence (KITCHENHAM; CHARTERS, 2007), (PETTICREW; ROBERTS, 2008).

We selected this type of literature review approach because our goal was to summarize the technology regarding PHRs and identify promising directions, which do not require an in-depth analysis and synthesis. With this in mind, we followed widely recognized empirical guidelines (KITCHENHAM; CHARTERS, 2007), (PETTICREW; ROBERTS, 2008) to plan and run systematic mapping studies. Moreover, to mitigate threats to validity, we followed the well-documented study protocol available in the studies by Biolchini et al. (BIOLCHINI et al., 2005) and Qiu et al. (QIU et al., 2015). The presented SLR method was carried out by defining the following activities:

(a) Research questions — introduce the research questions investigated;

(b) Search strategy — outline the strategy and libraries explored to collect data;

(c) Article selection — explain the criteria for selecting the studies;

(d) Distribution of studies — present how studies are distributed chronologically;

(e) Quality assessment — describe the quality assessment of the selected studies;

(f) Data extraction — compare the selected studies and research questions.

The following sections describe how this process of mapping the study was carried out.

3.1.2 Research Questions

According to Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007) and Petticrew and Roberts (PETTICREW; ROBERTS, 2008), the definition of research questions is the most important part of any systematic review. Therefore, we seek to identify and classify the technology related to PHRs; the features, problems, challenges, and solutions that are currently being considered; and the research opportunities that exist or are emerging. In this sense, we have defined general and specific research questions. The general research questions have been refined into more specific questions to provide a thorough classification and thematic analysis, as well as to pinpoint promising research directions for further investigation. Our research questions are classified into two categories: General Question (GQ) and Specific Question (SQ). Table 2 lists all the research questions investigated.

Table 2 – Research questions.

Group and identifier	**Issue**
General Questions (GQ)	
GQ1	How would the taxonomy for PHR classification appear?
GQ2	What are the challenges and open questions related to PHRs?
Specific Questions (SQ)	
SQ1	What are the data types that are included in a PHR?
SQ2	What are the standards that apply to PHRs?
SQ3	What are the user types and profiles that interact with a PHR?
SQ4	What are the interaction types of a patient with a PHR?
SQ5	Which are the techniques used to input information into a PHR?
SQ6	What are the goals of a PHR?
SQ7	What are the types or models of architecture of PHRs?

Source: Prepared by the author

The GQ group of research questions concerns a broader classification and some challenges concerning PHRs. GQ1 refers to the question of classifying and defining the taxonomy for PHRs. This research question focuses on the interoperability capacity that a PHR can have. This question highlights the integration issues of a PHR that is created and maintained by systems that are developed using heterogeneous technologies. GQ2 refers to the key challenges and issues in using PHRs. This is the main factor that will serve as a direct influence in the PHR survey. The purpose is to identify the types of issues that have been raised in the literature in the last decade. The research focuses on identifying the main problems affecting the spread of PHR adoption by patients and health care providers. For this question, we are able to reason with regard to the issues and factors that consequently influence PHR adoption. With the general research questions, we have also explored some derived specific research questions (SQ group) to improve the study filtering process. These questions have been proposed to pinpoint questions surrounding the adoption of the PHR. SQ1 seeks to identify the data types that a PHR can contain. SQ2 investigates the types and profiles of users who interact with a PHR. SQ3 examines the types of standards that are used in PHR implementations. SQ4 seeks to show the interaction types that a patient has with a PHR. SQ5 concentrates on evaluating the techniques or methods used to input data into a PHR. SQ6 investigates the purposes of a PHR. Finally, SQ7 concentrates on the types and models of PHR architecture.

3.1.3 Search Strategy

The next step was to find a complete set of studies related to the research questions. This process involved the designation of search keywords and the definition of search scope (PETTICREW; ROBERTS, 2008). In the construction of search keywords phase, we defined keywords to obtain accurate search results. In their report, Kitchenham and Charters (KITCHENHAM; CHARTERS, 2007) suggest breaking down the research question into individual facets as research units, where their synonyms, acronyms, abbreviations, and alternative spellings are all included and combined by Boolean operators. In addition, Petticrew and Roberts (PETTICREW; ROBERTS, 2008) propose the PICOC (Population, Intervention, Comparison, Outcome, and Context) criteria, which can be seen as guidelines to define such research units. In focusing on defining the PHR technology, we defined broader PICOC criteria based on the general research questions. Our goal was to refine and answer the specific research questions, which are derived from the general research questions with a restricted focus. Therefore, under the PHR scenarios, we defined the PICOC criteria as follows.

3.1.3.1 Population

The populations involve keywords, related terms, variants, or the same meaning for the technologies and standards on PHRs. Therefore, the following search string in the textbox was defined for the selection:

> ((('personal' or 'patient' or 'private') and ('health') and ('record' or 'application' or 'management' or 'information')) or ('patient' and ('access' or 'portal')) or ('PHR'[1] or 'PHA'[2] or 'PHM'[3] or 'PHI'[4]))

[1] PHR: Personal Health Record;
[2] PHA: Patient Health Application;
[3] PHM: Private Health Management;
[4] PHI: Private Health Information.

3.1.3.2 Intervention

We used the following terms to better filter studies in line with the purposes: health data, health services monitoring and reporting, patient monitoring devices, remote health monitoring, and mobile health care devices.

3.1.3.3 Comparison

This case refers to the comparison of different architecture types and models of implementation of the PHR. In addition, we compared the different PHR types regarding coverage and localization.

3.1.3.4 Outcome

The outcomes related to factors of importance to practitioners (e.g., improved reliability) and, in particular, to the patient. With respect to PHRs, this might refer to reducing the cost of collecting data, improving health information quality, anticipating potential problems, and allowing the patients to interact with their health data.

3.1.3.5 Context

In this regard, we analyzed the context of PHR information coverage in terms of content such as standardization, information grouping, and security and privacy in the relationships between patients and health care providers. Hence, the final keyword set is displayed in the following textbox:

> Keywords = PICOC = Population AND Intervention AND
> Comparison AND Outcome AND Context

In the definition of search scope phase, the source studies were obtained from selected electronic databases by searching using the constructed research keywords.

3.1.4 Article Selection

Once we found all the related articles, we proceeded to remove the studies that were not as relevant and kept only those that were the most representative. Therefore, we removed the studies that did not address PHR specifically. To apply the exclusion criteria, we used the terms of population and intervention criteria as follows:

- Exclusion criterion 1 — the article does not address PHR or related acronyms (population criterion I).

- Exclusion criterion 2 — the article does not address “health data” or “health services” (intervention criterion II).

The steps of the filtering process are as follows: (1) impurity removal, (2) filter by title and abstract, (3) removal of duplicates, and (4) filter by full text. First, the impurities of the search results were removed. Some impurities, for example, the names of conferences correlated to the search keywords, were included in the search results because of the characteristics of the different electronic databases. Second, we analyzed the title and abstract of the articles and excluded those that did not address PHR as a subject. Third, all the remaining studies were grouped, and the duplicates were removed because some studies were in more than one database. Some studies remained that were not particularly related to this survey. We analyzed the full text to remove those that were not relevant.

3.1.5 Quality Assessment

Since it is important and essential to assess the quality of the selected studies, the quality criterion is intended to verify that the article is really a relevant study (KITCHENHAM; CHARTERS, 2007). We evaluated the selected articles with regard to the purpose of research, contextualization, literature review, related work, methodology, the results obtained, and the conclusion in accordance with objectives and indication of future studies. For this purpose, the quality was evaluated according to Table 3, where the questions to which the articles were submitted to validate that these studies met the quality criteria are listed.

Table 3 – Quality assessment criteria.

Id.	**Issue**
C1[1]	Does the article clearly show the purpose of the research?
C2	Does the article adequately describe the literature review, background or context?
C3	Does the article present the related work with regard to the main contribution?
C4	Does the article have an architecture proposal or research methodology described?
C5	Does the article have research results?
C6	Does the article present a conclusion related to the research objectives?
C7	Does the article recommend future works, improvements, or further studies?

[1] C: Criterion;
Source: Prepared by the author

3.1.6 Data Extraction

We also developed an evaluation form for the selected articles in order to gather information about the studies and the sections where we found answers to general and specific research questions, which are presented in Table 4. This table shows each item of the study related to the research question, allowing us to assess and extract details of the articles and understand how the studies have addressed the issues related to the proposed research questions. The aim was to direct the survey to specific points that would answer the research questions.

Table 4 – Review articles related to the research questions.

Section	Description	Research questions
Open content		
Title	Title of the scientific article.	GQ1[1], GQ2, SQ1[2], SQ2, SQ7.
Abstract	Summary of purpose, method and results.	GQ1, GQ2, SQ1, SQ2, SQ7.
Keywords	Words representing the text content.	GQ1, GQ2, SQ1, SQ2, SQ7.
Article content		
Introduction	Specifies the issue to be addressed.	All questions.
Background	Includes concepts related to the proposal.	All questions.
Method	Presents and describes the methodology.	All questions.
Results	Performs evaluation according to method.	All questions.
Discussion	Data quantified compared with the literature.	GQ2, SQ2-SQ7.
Conclusion	Findings related to the objectives.	GQ2, SQ2-SQ7.

[1] GQ: General Question; [2] SQ: Specific Question.

Source: Prepared by the author

3.2 Two-layer Architecture Proposal Method

This section presents specific aspects of methods used to support and evaluate the OmniPHR architecture proposal in two-layer format, as well as the methodology that we used to select related work regarding the architecture models.

3.2.1 Related Work Selection Method

Our initial basis for proposing the model comes from the analysis of recent works since 2012, from the publication of ISO/TR 14292 (ISO, 2012), considering the definition of PHR and EHR applied in architectures proposals. For the selection was defined the following search string, according to the nomenclature used in the ISO standards:

(('personal' or 'electronic') and ('health' or 'medical') and ('record' or 'records')) and
((('distributed' or 'decentralized') or ('client-server' or 'centralized')) and
('architecture' or 'model'))

These terms were applied in recognized research portals on the computing and health areas: ACM, Google Scholar, IEEE, PubMed, Science Direct, and Springer.

3.2.2 Evaluation Method

The scientific community has been using modeling and profiling methodology to evaluate mobile applications (BANERJEE; GUPTA, 2015) (AHMED et al., 2015). With this strategy, the goal is to describe and evaluate scenarios of use where OmniPHR can be applied. Following this methodology, Bossel (BOSSEL, 2013) defines five steps to carry out the process:

(i) Developing the Model Concept - This stage defines the purpose of the model, which is to represent a typical use.

(ii) Developing the Profiling Model - This phase describes the system states, which are the scenarios to which the model will be submitted.

(iii) Profiling of System Behavior - At this stage, the emphasis is on the behavior of the model.

(iv) Performance Evaluation, Policy Choice, and System Design - At this stage, the emphasis is on the choice of assessment criteria and policies.

(v) Mathematical Systems Analysis - At this stage, we performed the mathematical analysis of the results.

3.3 Performance Analysis Method

This section explains the methods used in the OmniPHR prototype, evaluation, and results collection. Due to the barriers to the adoption of distributed health records across different health providers and in accordance with the background that underlying PHR and Blockchain technology, we researched the state-of-the-art regarding open issues in this area. Below we explain how we researched and analyzed related work and then outline the steps used to evaluate the performance of the OmniPHR model. We first reviewed the state-of-the-art by analyzing articles related to OmniPHR, which implements Blockchain solutions applied to health records. For this review, we used strings combining the PHR and EHR definitions with Blockchain. We then submitted these strings to PubMed, Medline, CiteSeerX, Cochrane, HealthStar, Elsevier, and Google Scholar, which are common portals that index scientific studies in the area of Health and Information Technology. In addition to verifying the correct reunification of patients' scattered data, we evaluated non-functional requirements (GALSTER; BUCHERER, 2008; CHUNG et al., 2012). The requirements and statistical formulas used to collect the data are described below.

Initially, we counted the Mean Time Between Failures (*MTBF*):

$$MTBF = \frac{TotalWorkingTime - TotalBreakdownTime}{TotalBreakdownIncidences} \tag{3.1}$$

and Mean Time To Repair (*MTTR*):

$$MTTR = \frac{TotalBreakdownTime}{TotalBreakdownIncidences} \tag{3.2}$$

to compose the Availability (*A*):

$$A = \frac{MTBF}{MTBF + MTTR} \tag{3.3}$$

Finally, we evaluated the Performance (*P*) extraction arithmetic mean:

$$P = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{3.4}$$

through the accounting of main memory, storage occupation, response time and throughput, where *a* compose the values and *n* the total of observations.

3.4 Multi-Blockchain Proposal Method

In this section, we introduce the methodology we use to support our OmniPHR Multi-Blockchain proposal. The section explains the methods applied and the related work selected in the study (KOTHARI, 2004). As a methodology, we followed the seven steps described below (BASKERVILLE; PRIES-HEJE; VENABLE, 2009; SOHAIB et al., 2019):

(i) In this stage, we present the problematic and research questions that underlie the study, with the general methodology and related work, as well as background and terminologies.

(ii) We express the problems identified in the requirements that the solution should meet.

(iii) We model and present the architecture using design-thinking techniques.

(iv) We construct the prototype following the requirements specified in the previous steps.

(v) We perform the evaluation of the prototype, collecting and presenting the data obtained.

(vi) We explore the analysis of the data in relation to the feasibility of the project, discussing the results obtained, verifying the results in relation to the specified requirements, as well as comparing the results in relation to the related works.

(vii) This stage presents the conclusions about the actions to address the problematic identified in the first step, the limitations of the solution and address possible future studies.

As a methodology to select the related work in the literature review (BASKERVILLE; PRIES-HEJE; VENABLE, 2009), we elaborated a search string that could extract the main studies about the implementation and the challenges faced in the use of Blockchain technology for records of geographically distributed health systems. Hence, we elaborate the following search string:

Blockchain(s) + ((Distributed / Decentralized) +
Architecture) + ((Health(care) / Medical) + (Records / Data))

3.5 Semantic Interoperability Proposal Method

This section presents the methods used to support and evaluate the OmniPHR semantic interoperability proposal. Initially, as a way to direct this stage of the proposal, we formulated a General Research Question (GRQ), which we aim to answer: *How a model proposal for PHR can promote semantic interoperability between open and proprietary health standards?*

In addition, we have formulated two Specific Research Questions (SRQ) about our proposal, as explained below:

- SRQ.1: *How does the model proposal aim to obtain a unified and up-to-date PHR viewpoint for both patients and health care providers?*
- SRQ.2: *What other advantages patients and health care providers could have with the interoperability of health data?*

3.5.1 Related Work Selection Method

In order to reach the related work at this step of the research, we used recognized research portals by the scientific community about articles related to health informatics, as follows: ACM Digital Library, Bentham Science Publishers B.V., BioMed Central, Google Scholar, IEEE Xplore Digital Library, Journal of Medical Internet Research (JMIR), Medical Library Association, Oxford University Press, PubMed, SciELO, ScienceDirect (Elsevier BV) and SpringerLink. The research was limited to the search of articles written in the English language and from the last six years since the original PHR definition by ISO is from 2012. The search used the following initial search string to restrict and localize studies that specifically address the article keywords, where the asterisk symbol (*) means AND and sum symbol (+) means OR:

('personal health record(s)' + PHR) * (interoperability + inter-operability) * semantic *
(health * (standards + standardization))

3.5.2 Evaluation Method

After to select the related work, the next step of the methodology is to define in the OmniPHR proposal how the model can address related issues regarding semantic interoperability. For this purpose, we followed the same steps of the methodology proposed by Bossel (BOSSEL, 2013) that we used to evaluate the architecture proposal. In addition, to evaluate the OmniPHR model specifically related semantic interoperability, we proposed to use a real database of patients' health anonymized data. In this way, we can to compose and submit different types of health data standards to the model. After evaluation, the next step is to analyzes and discusses the results obtained, presenting conclusions about the findings. Regarding the evaluation, the proposal is to evaluate the model obtaining a statistical analysis of the solution. In this sense, a metric recognized by the scientific community is the F1-score (or F-measure) (POWERS, 2011). Following this metric, the precision and recall of the algorithm are calculated. We also calculated the accuracy, to have a measure in relation to the total records.

The precision (or ppv = positive predictive value, also known as $confidence$) is given by

$$ppv = \frac{tp}{tp + fp} \tag{3.5}$$

The recall (or tpr = true positive rate, also known as $sensitivity$) is given by

$$tpr = \frac{tp}{tp + fn} \tag{3.6}$$

And accuracy (acc, also known as $trueness$) is given by

$$acc = \frac{tp + tn}{tp + tn + fp + fn} \tag{3.7}$$

Where tp = true positives, tn = true negatives, fp = false positives, and fn = false negatives. The F1-score is defined as follows:

$$F1 = 2 \cdot \frac{ppv \cdot tpr}{ppv + tpr} \tag{3.8}$$

Applied to this proposal, accuracy represents the proximity measure between the number of converted and unconverted fields, as expected, in relation to the total measured fields. Precision represents the number of fields converted correctly, as expected, divided by the number of fields returned in the process execution. The recall represents the number of successfully converted fields, as expected, divided by the number of fields that should have been converted. The harmonic mean of precision and recall results in the F1-score.

4 RELATED WORK

This chapter concentrates the main related work in order to support the proposed solution, following the methodology previously presented. We published a full article (ROEHRS et al., 2017) contemplating the related work criterion in the Journal of Medical Internet Research (JMIR). The chapter is divided into six sections.

(i) Section 4.1 has a higher level of detail since the section presents the step-by-step of SLR methodology development regarding the state-of-the-art of PHR. This section also presents findings regarding the PHR review.

(ii) Section 4.2 presents the related work selected to support the architecture model.

(iii) Section 4.3 presents the related work to support the performance analysis of the model.

(iv) Section 4.4 presents the related work regarding the Multi-Blockchain proposal.

(v) Section 4.5 presents the related work selected to support the OmniPHR semantic interoperability proposal.

(vi) Section 4.6 presents research opportunities collected from related work.

4.1 Related Work on Systematic Literature Review

4.1.1 Recruitment

In this section, we present the results obtained from the 48 fully assessed studies related to the research topic. We seek to answer each proposed research question in the following subsections through elaborative information synthesis. As a result, aside from answering the research questions, we have also proposed contributions in the PHR field from the study of related works, which are an updated taxonomy and an updated vision about main challenges and issues, as well as an updated survey about data types, standards, user types, profiles, and input techniques.

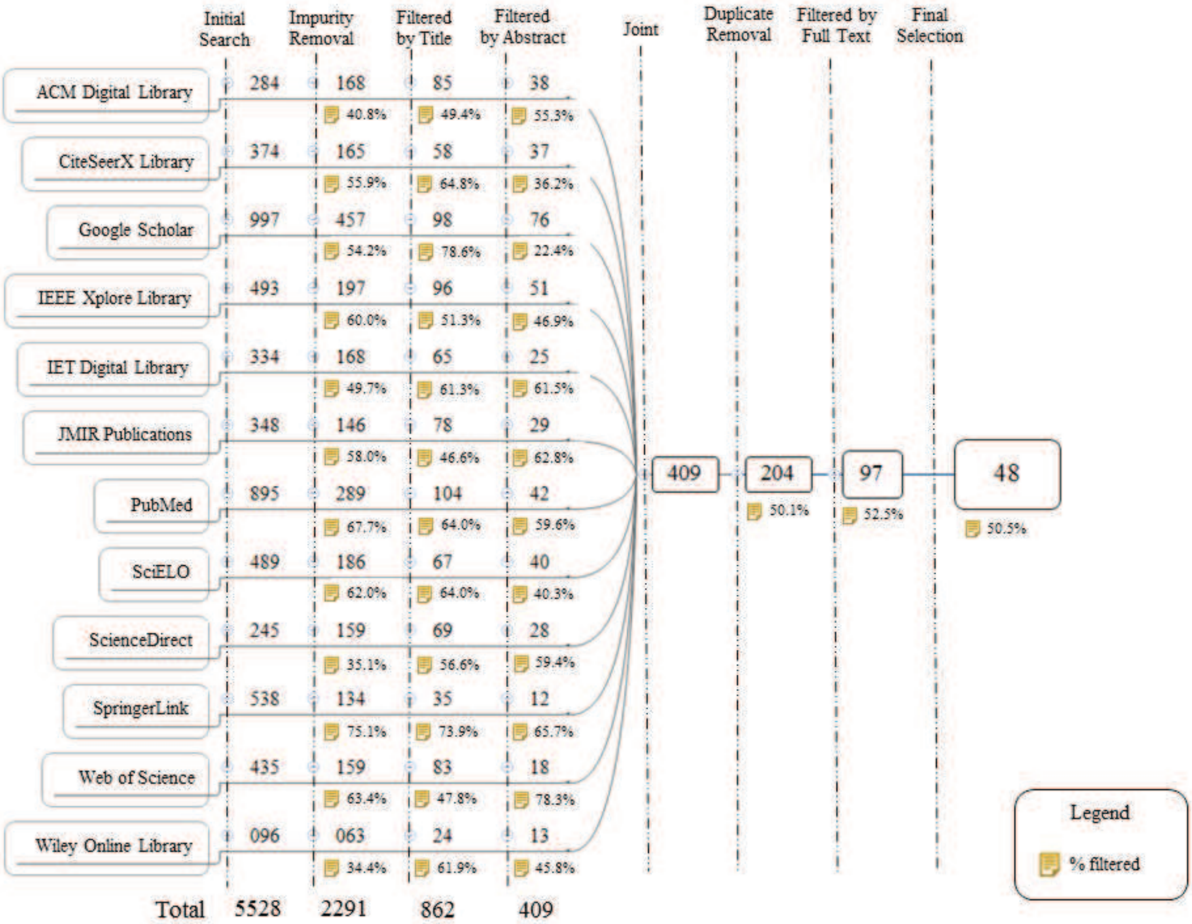
4.1.2 Conducting the Search Strategy

To cover as many related studies as possible, we selected 12 electronic databases as our search scope, which are listed in Appendix 1. These portals cover the most relevant journals and conferences within the computer science and health care field. In Appendix 2, we present the publishers or organization editors and the respective publications of the selected studies. Duplicated results produced from different databases were excluded by manual filtering in the study selection.

4.1.3 Proceeding with Article Selection

The selection process is summarized in Figure 3, which shows the filtering process. We found 5528 articles in the initial search before applying the exclusion criteria; of these, 3237 (58.55%) articles were identified as impurities. We applied the first exclusion criterion to the studies that remained after we withdrew these articles. Continuing the process, 1429/2291 (62.37%) articles were filtered through a title review, and 453/862 (52.5%) articles were filtered through abstract analysis. We grouped the studies that remained, and 205/409 (50.1%) articles were identified as duplicates and were removed. After this stage, exclusion criterion 2 was applied to the full text and only 97/204 (47.5%) remained.

Figure 3 – Systematic mapping study – article selection.



Source: Prepared by the author.

When analyzing the 97 candidate articles in the list, we noticed that some of these studies were from the same author or research group and were similar in many respects. For articles that were repeated, the most representative article was selected. Thus, 49/97 (50%) articles were excluded at this stage. Finally, 48 articles were selected as the baseline for the study. An overview of all primary studies is presented in Table 5 with the identifier, reference, publication year, publisher, and type, which are sorted in ascending order by publication year.

Table 5 – List of articles for SLR.

Identifier	Study, year	Publisher	Type
A01	(BRICON-SOUF; NEWMAN, 2006)	Elsevier	Journal
A02	(TANG et al., 2006)	Oxford[1]	Journal
A03	(FROST; MASSAGLI, 2008)	JMIR[2]	Journal
A04	(KAELBER et al., 2008)	Oxford	Journal
A05	(HUDA; YAMADA; SONEHARA, 2009)	IEEE[3]	Conference
A06	(KIM et al., 2009)	JMIR	Journal
A07	(BRENNAN; DOWNS; CASPER, 2010)	Elsevier	Journal
A08	(CASTILLO; MARTÍNEZ-GARCÍA; PULIDO, 2010)	BioMed[4]	Journal
A09	(HORAN; BOTTS; BURKHARD, 2010)	JMIR	Journal
A10	(HUDSON; COHEN, 2010)	IEEE	Conference
A11	(JONES et al., 2010)	MLA[5]	Journal
A12	(NAZI et al., 2010)	Springer	Journal
A13	(PATEL et al., 2010)	Elsevier	Journal
A14	(RETI et al., 2010)	Oxford	Journal
A15	(WEN et al., 2010)	JMIR	Journal
A16	(WILLIAMS, 2010)	ACM[6]	Conference
A17	(WYNIA; DUNN, 2010)	Wiley	Journal
A18	(ARCHER et al., 2011)	Oxford	Journal
A19	(BAIRD; NORTH; RAGHU, 2011)	ACM	Conference
A20	(CALIGTAN; DYKES, 2011)	Elsevier	Conference
A21	(LAFKY; HORAN, 2011)	SAGE	Journal
A22	(LIU; SHIH; HAYES, 2011)	ACM	Conference
A23	(SIEK et al., 2011)	Springer	Journal
A24	(ZULMAN et al., 2011)	ACP[7]	Journal
A25	(SEÑOR; FERNÁNDEZ-ALEMÁN; TOVAL, 2012)	JMIR	Journal
A26	(EMANI et al., 2012)	JMIR	Journal
A27	(FUJI et al., 2012)	Springer	Journal
A28	(KHARRAZI et al., 2012)	Elsevier	Journal
A29	(LUO; TANG; THOMAS, 2012)	Springer	Journal
A30	(STEELE; MIN; LO, 2012)	Wiley	Journal
A31	(SUNYAEV; CHORNYI, 2012)	ACM	Journal
A32	(AGARWAL et al., 2013)	JMIR	Journal
A33	(LI et al., 2013)	IEEE	Journal
A34	(NAZI, 2013)	JMIR	Journal
A35	(WOODS et al., 2013)	JMIR	Journal
A36	(ANCKER; SILVER; KAUSHAL, 2014)	Springer	Journal
A37	(BOURI; RAVI, 2014)	JMIR	Journal
A38	(CAHILL; GILBERT; ARMSTRONG, 2014)	Springer	Journal
A39	(CHRISCHILLES et al., 2014)	Oxford	Journal
A40	(OZOK et al., 2014)	Elsevier	Journal
A41	(SPIL; KLEIN, 2014)	IEEE	Conference
A42	(WELLS et al., 2014)	Oxford	Journal
A43	(CZAJA et al., 2015)	SAGE	Journal
A44	(LIU; HUANG; LIU, 2015)	Elsevier	Journal
A45	(PRICE et al., 2015)	BioMed	Journal
A46	(SPIL; KLEIN, 2015)	Elsevier	Journal
A47	(SUJANSKY; KUNZ, 2015)	Springer	Journal
A48	(FORD; HESSE; HUERTA, 2016)	JMIR	Journal

[1] Oxford: Oxford University Press; [2] JMIR: JMIR Publications;

[3] IEEE: Institute of Electrical and Electronics Engineers; [4] BioMed: BioMed Central;

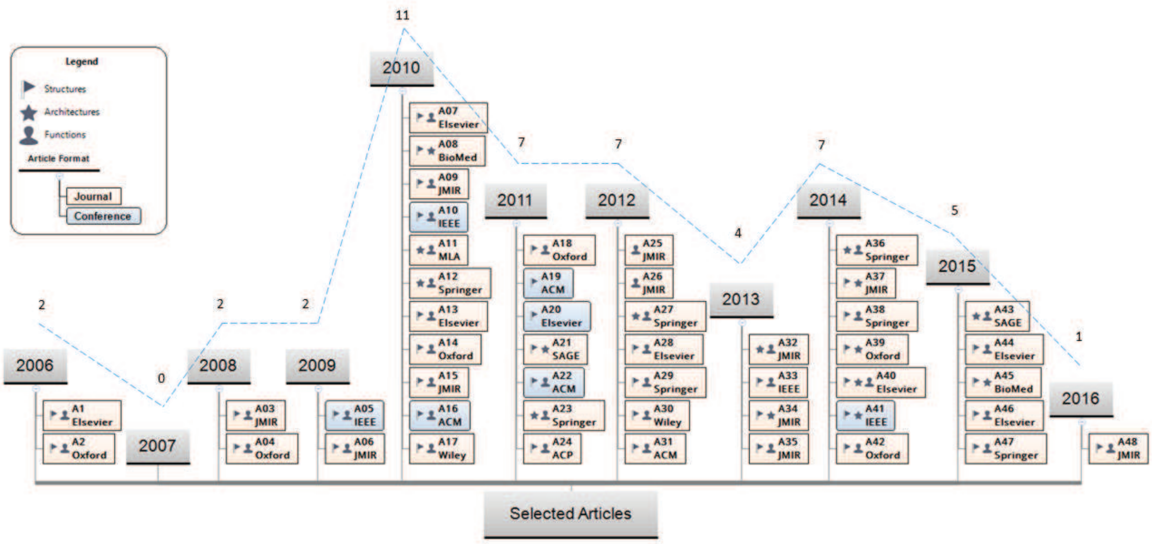
[5] MLA: Medical Library Association; [6] ACM: Association for Computing Machinery;

[7] ACP: American College of Physicians.

Source: Prepared by the author

In Figure 4, we present the evolution of the selected publications over the years, ranging from 2006 to 2016. The studies were analyzed according to the main objectives, as seen in the figure legend, where the articles were divided into the groups “Structures,” “Architectures,” and “Functions.” Above each year, the number of articles published in that year is shown. Each item label includes the publisher of the work, and the journal and conference articles are distinguished by the box format.

Figure 4 – Publication chronology.



Source: Prepared by the author.

4.1.4 Performing the Quality Assessment

In Figure 5, we present the quality criteria score of the articles based on the quality assessment criteria proposed in Table 2. The quality criteria score each article obtained is shown on the vertical axis and the studies themselves on the horizontal axis, from 1 to 48. Upon analysis, most articles met all the criteria for evaluation, responding positively to at least 6 out of 7 quality assessment criteria. For instance, several articles do not comment on or cite possible future studies in general because they are conclusive articles, with a conclusion on its assessment.

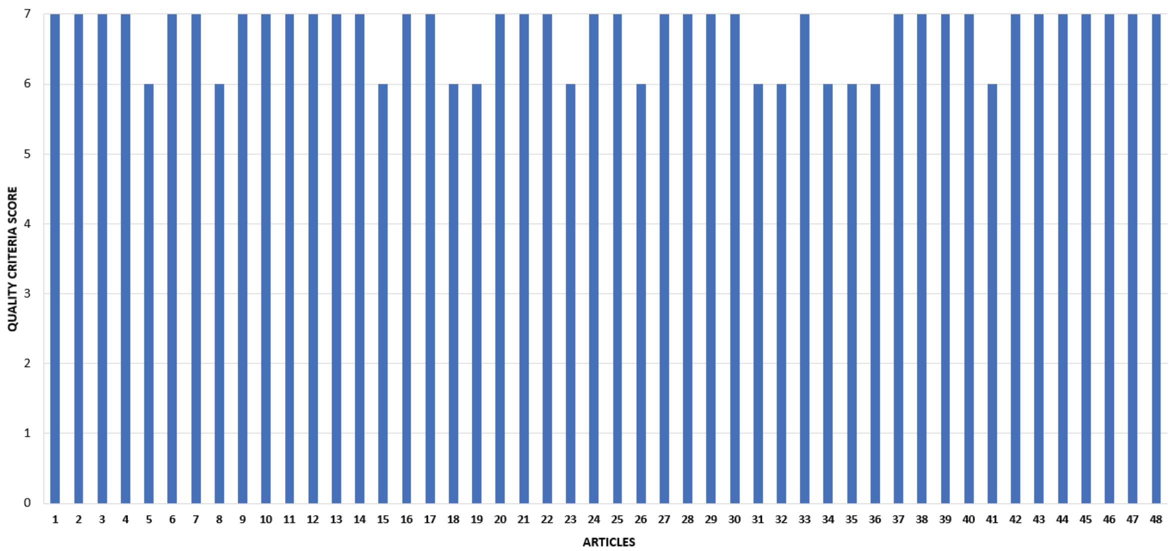
4.1.5 Data Extraction and Answers to the Research Questions

Finally, to address the general research questions, we have identified the following.

4.1.5.1 GQ1: How Would the Taxonomy for PHR Classification Appear?

We identified studies that investigated a number of current issues that were addressed in the PHR field. Therefore, we managed to build the proposed taxonomy to gather and organize

Figure 5 – Quality assessment of the articles.



Source: Prepared by the author.

the various possibilities for PHRs. By analyzing the selected articles and seeking to answer this general research question, we propose a taxonomy for PHR based on important characteristics of the models, and we believe that this taxonomy could help to classify, compare, and evaluate different PHR types. Moreover, this classification can provide an overview of possible alternatives in terms of aims, content, and architectures. The proposed taxonomy for the PHR classification is summarized in Table 6, which is broadly divided into three groups: (1) Structures, (2) Functions, and (3) Architectures. The specific research questions (SQ1 to SQ7) are included in the taxonomy, which was developed through analysis of the selected articles.

Table 6 – Personal Health Record taxonomy.

Group and item	Description
Structures	Main data types and standards used in health records.
Data types	Data types found in PHRs[1] (see subsection SQ1[2]).
Standards	Standards to which PHRs can adhere (see subsection SQ2).
Functions	Depicts the main goals and features present in the PHRs.
Users profiles	User types and profiles that interact (see subsection SQ3).
Interaction	Patient's interaction types with a PHR (see subsection SQ4).
Data source	Techniques for input of information (see subsection SQ5).
Goals	Represents the aim of the PHR (see subsection SQ6).
Architectures	Architecture types and scopes (see subsection SQ7).
Models	Describes the main architecture models.
Coverage	Has a physical location division for data.

[1] PHR: Personal Health Record; [2] SQ: Specific Question.
Source: Prepared by the author

4.1.5.2 GQ2: What Are the Challenges and Open Questions Related to PHRs?

To answer this question, we listed and identified challenges, open questions, aspects, issues, and common concerns in the adoption of PHR among the analyzed studies. These aspects were collected and are presented in Table 7. As seen, the content is split to group some of the common characteristics of challenges and concerns (GCC – group of challenges and concerns) related to collaboration and communication (GCC1), privacy, security, and trust (GCC2), infrastructure (GCC3), and integration (GCC4). The subject most cited ranges from CC01 to CC15.

Table 7 – Personal Health Record challenges and concerns.

Group / ident.	Challenge and concern	Reference articles
GCC1[1]: collaboration and communication		
CC01[2]	Context-aware computing	A01, A41.
CC02	Wearable computing, IoT[3]	A01, A28.
CC03	AI[4] applied to health	A01, A10, A16.
CC04	Personaliation, usability, familiarity, comfort	A02, A07, A19, A22, A29, A40, A42, A45.
CC05	Manage medications	A23, A29.
CC06	Patient-generated data	A22, A42, A44, A45, A47.
GCC2: privacy, security, and trust		
CC07	Confidentiality and integrity	A07, A08, A19, A29, A42, A45, A46.
CC08	Data repository ownership	A13, A16, A19, A45, A47.
CC09	Authorization and access control technologies	A02, A07, A11, A16, A21, A22, A31, A40, A42.
CC10	Secure transport protocol	A16, A22, A42, A47.
GCC3: infrastructure		
CC11	Portability - devices, equipment, HW[5]	A11, A18, A21, A23, A24, A28, A30, A42, A43, A44.
CC12	Efficiency and scalability	A01, A40, A41, A44, A45, A46.
GCC4: integration		
CC13	Patterns in collecting medical data	A13, A17, A42, A47.
CC14	Terminology	A22, A29.
CC15	Interoperability	A13, A16, A21.

[1] GCC: Group of Challenges and Concerns; [2] CC: Challenge and Concern; [3] IoT: Internet of Things;
[4] AI: Artificial Intelligence; [5] HW: Hardware.
Source: Prepared by the author

In GCC1 group, there are challenges and issues related to collaboration and communication, ranging from data types to be stored and made available in the PHR to policy barriers to limit the provided information type. Some articles mention the PHR data that are available according to the context awareness, such as CC01, and some articles discuss wearable computing and IoT, such as CC02. Other articles examine AI that is applied to the health sector in CC03. The customization, usability, familiarity, and comfort when using the PHR is the subject matter of several articles in CC04, and the management of medications contained in the PHR is reviewed in CC05. GCC2 group presents issues related to privacy, security, and reliability that are presented in PHRs: CC07 addresses confidentiality and integrity issues. CC08 refers to data repositories and their owners. CC09 examines access control technologies. CC10 includes a discussion on data transport protocols. GCC3 group treats issues related to the infrastructure of PHRs, in which CC11 discusses the portability of devices and equipment used with a PHR. In CC12, issues on the efficient construction of computer systems and the scalability of the infrastructure used to support PHR solutions are discussed. Finally, in the GCC4 group, concerns about integration are examined, such as in CC13, which concerns patterns in collecting medical data. CC14 presents concerns about the terminology used to collect and store PHRs. CC15 addresses issues about interoperability.

Regarding specific research questions, we have identified the following:

4.1.5.3 SQ1: What Are the Data Types That Are Included in a PHR?

To answer this research question, we analyzed all selected studies that involved research of the data types used in PHRs, which are summarized in Table 8. Through the analysis of proposals and references in selected articles, we were able to obtain an updated set of data types related to PHRs. The data types ranged from information cited in many studies, such as those on allergies, immunizations, and medications, to types that are not frequently mentioned, such as genetic information and home monitoring data.

Table 8 – Personal Health Record data types.

Type	Description	Reference articles
Allergies	Allergies and adverse reactions	A02, A12, A16, A18, A20, A25, A28, A30, A35, A39, A40, A41, A46.
Demographic	Patient statistics and clinical data	A03, A20, A35, A39, A40, A43.
Documents	Attached files (photos, scanned documents)	A07, A20, A28.
Evolution	Progress and clinic notes, care plan	A07, A14, A18, A34.
Family history	Family medical history	A02, A12, A16, A18, A20, A25, A28, A37.
General	Patient registration info., emergency contact	A03, A12, A16, A18, A28.
Genetic	Genetic information	A16, A25.
Home monitor	Home-monitored data	A02, A18, A25.
Immunizations	Immunization records (vaccine), tracking immunizations	A02, A09, A12, A16, A18, A19, A20, A25, A28, A30, A32, A37.
Insurance	Insurance plan information, billing	A16, A18, A28.
Laboratory results	Laboratory and imaging test results (laboratory tests)	A02, A12, A14, A16, A18, A19, A20, A25, A28, A32, A35, A43.
Major illnesses	List of major diseases	A03, A02, A12, A18, A25.
Medications	Medication list prescribed, past medicines taken	A02, A07, A12, A16, A18, A20, A25, A28, A35, A39, A41.
Prescriptions	Medical prescription refills (renewing)	A04, A09, A12, A15, A17, A43, A46.
Prevention	Preventive health recommendations	A12, A18, A32, A40, A46.
Providers	Previous health care provider list	A02, A18, A28, A30, A37.
Scheduling	Appointments, past procedures, hospitalizations	A02, A12, A16, A18, A20, A25, A28, A35, A37.
Social history	Social history, lifestyle (health habits)	A02, A12, A18, A25, A40.
Summaries	Admissions, permanencies, and discharges	A39, A35, A43.
Vital signs	Status of bodily functions	A16, A30, A35, A37, A40.

Source: Prepared by the author

4.1.5.4 SQ2: What Are the Standards That Apply to PHRs?

Some providers use proprietary formats to organize their health records that are used only by internal applications, each of which has a different format (WYNIA; DUNN, 2010), (SUJANSKY; KUNZ, 2015). Thus, to answer this question, we focused on open standards, which are summarized in Table 9 and present a vast number of organizational data patterns for health records. Table 9 lists the referenced standards (GS – group of standards) according to their goals: nomenclature and terminology (GS1), privacy (GS2), structural and semantic (GS3), and templates and technology platforms (GS4). In group GS1, standards regarding nomenclature and terminology were grouped. Group GS2 contains only one standard that addresses privacy. In the GS3 group, several structural and semantic standards are presented. Finally, the GS4 group is related to templates and technology platform standards. We were able to identify some standards from the research on integrations and related projects, such as *open*EHR (OPENEHR, 2017), which is integrated with the DICOM standard and others.

Table 9 – Main personal health record–related standards.

Group and standard	Description	Reference articles
GS1[1]: nomenclature and terminology		
HNA/NIC[2]	Classifications of nursing activities and interventions.	A29.
ICDx	Family of international classification of diseases.	A11, A28, A29, A44.
LOINC	Code names for identifying medical observations.	A47.
SNOMED CT	Terminology collection of medical terms.	A11, A28, A47.
UMLS	System of medical vocabularies.	A11, A13.
GS2: privacy		
HIPAA	USA legislation for medical information.	A09, A22, A25, A35.
GS3: structural and semantic		
ASC X12N	Accredited standards committee X12-INS.	A45, A47.
CCD	Specification for exchange clinical documents.	A11, A47, A48.
CCR	Specification for sharing continuity of care content.	A11, A33.
CDA	Specification for clinical notes.	A11, A47.
DICOM	Standard for medical digital imaging.	A11.
EN 13606	EHR[3] standards in Europe.	A25.
HL7/FHIR/ SMART	Family of standards and platforms based on the HL7 reference model.	A11, A18, A28, A42, A43, A45, A47.
ISO[4]	TR (Technical Report) 14292 (PHR) and ISO/IEEE 11073 Personal Health Data (PHD).	A01, A03, A20, A23, A25, A38, A43, A47.
*open*EHR	Open standards specification in eHealth.	A11.
xDT	German family of data exchange formats.	A04.
GS4: templates and technology platforms		
OpenMRS	Platform and reference application named Open Medical Record System.	A42.
OSCAR	EHR system named Open Source Clinical Application and Resource.	A42.

[1] GS: Group of Standards; [2] HNA/NIC: Home Nursing Activities/Nursing Interventions Classification;
[3] EHR: Electronic Health Record; [4] ISO: International Organization for Standardization.

Source: Prepared by the author

4.1.5.5 SQ3: What Are the User Types and Profiles That Interact With a PHR?

Upon analyzing the selected articles, we identified a set of profiles or user types that have access to the electronic patient record, which vary from the physician, who is primarily responsible for the PHR information, to the patient. The types of access also include the possibility that some data may be publicly available, for example, on social networks (LUO; TANG; THOMAS, 2012). There are multiple stakeholders involved in accessing the PHR, such as patients, providers, employers, payers, governments, and research institutions (TANG et al., 2006). In Multimedia Appendix 3, we present the details of the profiles that have been identified. We can see that the physician is widely referenced, while the nurse and administrative profiles are not cited as often. Among the laity, the patient profile is often cited; however, the relative or guardian profile is less commonly cited. We also included a public profile because patients might share their information anonymously in some cases or for other cases in which public administration sectors provide open statistical data.

The following is a brief description of perceived profiles:

(a) Physician or doctor — the physician, in this assessment, is the health professional profile responsible for reporting patient data in consumer electronic records.

(b) Nurse — according to the International Standard Classification of Occupations (ISCO, 2016), nursing professionals provide treatment, support, and care for people who need nursing care owing to the effects of aging, injury, disease, or other physical or mental impairments or face potential risks to their health.

(c) Administrative — this profile refers to all administrative health professionals who are not directly linked to the data generation but have informational access for bureaucratic, statistical data gathering, or financial information needs.

(d) Patient or consumer — this profile refers to the PHR principles; some authors also refer to the patient as a consumer of health care (LAFKY; HORAN, 2011), (CALIGTAN; DYKES, 2011).

(e) Relative — this profile is composed of parents, guardians, caregivers, responsible legal individuals, or anyone who has the patient's permission to access his or her PHR.

(f) Public or anonymous — this refers to profiles with external access in an anonymous or public way, such as institutions, the government, researchers, health plans, third parties, and even social networks.

4.1.5.6 SQ4: What Are the Interaction Types of a Patient With a PHR?

This research question seeks to describe the interaction types of a patient with a PHR, that is, the types of relationships that a patient has using the PHR. In the following section, we present a brief description of the interaction types that were identified when analyzing the articles:

(a) Direct — in this case, the patients are the owners and manage their health data in the PHR. Reference articles: A02, A05, A09, A12, A25, A26, A31, A48.

(b) Indirect — in this case, the patient has read-only access and cannot edit the data. The health care providers are the owners, and the patient can only download or print the health records. Reference articles: A01, A05, A22, A25, A26, A40, A41, A42.

(c) Outsourced — in this case, the patient authorizes a third party to handle the health data or the responsible parties (e.g., parents) manage the patient's health records. Reference articles: A02, A03, A04, A07, A18, A24, A25, A28, A37, A48.

4.1.5.7 SQ5: Which Are the Techniques or Methods Used to Input Information Into a PHR?

Another result was the identification of techniques and actors that interact in the process of data collection for inputting into a PHR. Table 10 presents some answers to this specific research question, summarizing the techniques of inputting the relevant data into PHRs. This information follows standards and is intended to structure and standardize the data provided. We list the main profiles (actors) that provide the data, including health professionals and the patients themselves, which are gathered from the environment, including anonymously. The techniques (T) identified for inputting data range from data collaboration (T1), to patient reports (T2), adaptive platforms (T3), and anonymization (T4). Table 10 also includes articles in which these techniques and actors are cited. In short, this was the actors' group that was identified with a relevant interaction in collecting data for inputting data into the PHR.

Table 10 – Techniques for inputting information into personal health records.

Techniques and profiles	Description	References
Data collaboration (T1[1])		
Health professionals	Collaboration between multiple health care professionals. Health care providers are the owners (paternalistic relationship).	A08, A09, A12, A15, A22, A23.
Patient reports (T2)		
Patient	Patient reports data, for example, listing drugs that are being used or menstrual period data.	A23, A26, A47.
Adaptive platforms (T3)		
Environment	Aggregate sources provisioning individualized personal eHealth services combined with context information, including monitoring sensors. Patient and health care providers collaborate for inputting data into PHR[2].	A01, A26, A38, A43, A44.
Anonymization (T4)		
Anonymous	Anonymizing social network data.	A16, A44.

[1] T: Technique; [2] PHR: Personal Health Record.

Source: Prepared by the author

4.1.5.8 SQ6: What Are the Goals of a PHR?

This research question includes the main goals of the PHR. This question is intended to identify the purpose that a PHR has in a broad context, and that applies to any profile that has access. In the following section, we present a brief description of the interaction types:

(a) Consult — in this case, the purpose is to allow the profile only to consult (in read-only mode). Reference articles: A01, A03, A07, A10, A13, A15, A16, A17, A21, A39, A47.

(b) Maintain — in this case, the user profile is allowed to maintain and control the health records. Reference articles: A09, A16, A18, A22, A29, A33, A37, A46.

(c) Monitor — in this case, the PHR is in monitoring mode and can send alerts or warnings for one or more profiles; the goal is to help the patients monitor their health. Reference articles: A01, A07, A10, A20, A23, A25, A29, A40, A43, A45.

4.1.5.9 SQ7: What Are the Types or Models of Architecture of PHRs?

The purpose of this question is to identify the types or models of architecture in which a PHR can be implemented. When analyzing the articles, as seen in Table 11, the architecture types (AG – architecture group) were split into two groups: model (AG1) and coverage (AG2). The first group, AG1, describes the main architecture models. The second group, AG2, divides the data based on the physical location, that is, the scope of the PHR.

Table 11 – Personal Health Record architecture types or models.

Group and item	Description	Reference articles
AG1[1]: model		
On paper	Health records are kept on paper.	A08, A20, A22.
Inside	PHR[2] is kept in local repositories, inside the provider, for example.	A02, A03, A16, A20, A31.
Outside	PHR is distributed or shared between servers outside the provider.	A01, A03, A24, A35.
Hybrid	PHR is distributed inside and outside the provider.	A02, A10, A28, A35, A47.
AG2: coverage		
Stand-alone	Data coverage is used only in the provider area.	A11, A26, A45, A46.
Local	Area is at the city level.	A03, A11, A20, A29, A35.
Regional	Data are used in the state or province.	A02, A04, A25, A37, A45.
National	Coverage encompasses the nation.	A09, A12, A28, A34, A35.
International	Coverage transcends the nation.	A09, A16, A28, A30.

[1] AG: Architecture Group; [2] PHR: Personal Health Record.

Source: Prepared by the author

4.1.6 Systematic Literature Review Findings

In this study, we sought to identify a quantitative and qualitative sample of studies that enabled us to obtain a clear overview of the technology regarding PHRs in the last 10 years from a number of candidate articles. This research sought to highlight some of the most relevant studies of the field according to certain systematic selection criteria. The survey sought to identify several common aspects of studies by answering a number of research questions. As a result, we were able to propose a PHR taxonomy and identify gaps to be further researched that represent challenges and issues that have been detected in recent years. These aspects range from patients' concerns to providers' problems regarding PHR adoption. In addition, we have identified the data types included in PHRs, an updated tabulation of the data standardization, access profiles, and their characteristics, and, finally, a classification of input techniques. We also identified other common and related aspects. These opportunities are discussed as follows.

4.1.6.1 GQ1: How Would the Taxonomy for PHR Classification Appear?

For the GQ1 research question, we sought to define a PHR taxonomy, which is presented in Table 5. Our proposed taxonomy illustrates the PHR types and their organization according to several studies that were analyzed. We primarily identified three major groups of PHR organization types: (1) Structures, (2) Functions, and (3) Architectures. From these groups, we were able to examine the PHR types in-depth to understand each one of them. These groups also showed that there are PHR application initiatives on several fronts with concerns that range from features and content to architectural format in terms of PHR implementation (STEELE; MIN; LO, 2012).

4.1.6.2 GQ2: What Are the Challenges and Open Questions Related to PHRs?

For the GQ2 research question, we sought to define the main challenges and issues regarding the use of PHRs. There are many open questions to be further researched in the area of PHR. The challenges and constraints in the adoption of PHRs are diverse. Some research results indicate problems of usability, privacy, security, and complexity in the use of PHRs, ranging from fears of including erroneous data to the difficulty of interpretation as the main difficulties (BAIRD; NORTH; RAGHU, 2011), (LIU; SHIH; HAYES, 2011). In Table 6, we describe some challenges and issues that may give rise to future studies. According to the number of items in each group in the table, we notice a greater concern with the first three groups, although we cannot claim this assessment as being definitive. One possibility that we touch upon for this observation is that the integration of standards and interoperability, as well as the nomenclatures and terminologies, are already in a stage of stability and consolidation. This leads us to reinforce the thesis that the concerns of the authors at this time are the issues raised by the first three groups of problems. That is, the concerns and challenges are more focused on discussions regarding confidentiality, integrity, authorization, access control, portability, efficiency, scalability of solutions, and issues related to user experience.

4.1.6.3 SQ1: What Are the Data Types That Are Included in a PHR?

With respect to the SQ1 research question, we sought to define an updated ranking on data types in PHRs. Upon analyzing the studies, we observed that PHR data types have evolved since the first PHRs (TANG et al., 2006), (BRICON-SOUF; NEWMAN, 2006). The data types found include groups that are not usually included in EHRs. Among the EHR stored data are medications, prescriptions, scheduled appointments, vital signs, medical history, laboratory information, immunizations, summaries, scanned documents, billing information, and progress notes about changes in the patient's health (ISO, 2005). However, in PHRs, new data types have emerged, including genetic information (SEÑOR; FERNÁNDEZ-ALEMÁN; TOVAL, 2012), (WILLIAMS, 2010), medical advice (recommendations), and prevention concerning the patient's health, as well as data types with recommendations for prevention and home monitoring data (SPIL; KLEIN, 2015), (OZOK et al., 2014). Other data types that appear in PHRs are allergies, patient registration data, and insurance plan information, including demographic data such as age, sex, and education. Furthermore, information on the patient's family, social history, lifestyle, food, diet, daily activities, and a list of providers who treated the patient previously are included in PHRs.

4.1.6.4 SQ2: What Are the Standards That Apply to PHRs?

For the SQ2 research question, we sought to define a current view of PHR standards. The result was the identification of the current list of existing data standards used in PHRs. We observed several standards that were maintained by various stakeholders that were located in different countries and regions. We were also able to observe a consolidation of some patterns in the articles' citations, such as ISO (ISO, 2005), (ISO, 2012) and HL7 (Health Level Seven) (ARCHER et al., 2011), which are used to define and establish interoperability between the systems. When analyzing the articles, it was observed that all the standards listed could be used directly or indirectly with a PHR. However, their purposes are diverse. Some standards have specific goals, for example, DICOM (JONES et al., 2010) and SNOMED CT (SUJANSKY; KUNZ, 2015), while others have broader purposes, for example, HL7 (ARCHER et al., 2011) and *open*EHR (OPENEHR, 2017), which can be integrated with other specific standards to render the solution. Finally, we identified some open systems or platforms that serve as templates, which use some of the listed standards to propose management solutions for patients' health data.

4.1.6.5 SQ3: What Are the User Types and Profiles That Interact With a PHR?

In the SQ3 research question, we sought to define the PHR user types and profiles that address PHR. The result was the identification of updated profiles as well as their characteristics. For the security and privacy of the health data, the answer to this research question offered a clear definition of the profiles that are allowed access to the PHR and what their responsibilities are (FUJI et al., 2012). In terms of access profiles, although the PHR is focused on personal use, the idea is that a patient can also delegate access to third parties by choice or necessity, as in the case of children or people who need special care. These third parties can access all or only specific parts of the PHR dataset. Patients can share their PHR for various purposes. Such patients may be minors whose parents need to share their health data with physicians, people with special needs who require constant monitoring, or even patients who wish to share their health data with other physicians.

By analyzing the selected articles, it was possible to find multiple profiles that have access to the PHR. Therefore, e can highlight the following profiles: patients, physicians, nurses, relatives, administrators, and the public. A physician's tasks include recording the health information and medical history of the patients as well as exchanging information with practitioners and other health care professionals (ISCO, 2016). In cases where patients need emergency care, a primary care physician usually treats them. If more specialized care is needed, the physician indicates the need for a specialist. Furthermore, physicians must report births, deaths, and notifiable diseases to the government. Because the PHR is composed of health data that are stored for a lifetime, many physicians edit the PHR over time. Otherwise, in the case of an administrative profile, these professionals usually have limited and controlled access to medical records. This profile is considered internal access, which is not to be confused with external access institutions.

With the patient profile, the user can manage the information provided in his or her repository. The purpose is for patients to have access to their health data and use them throughout their lives (SUJANSKY; KUNZ, 2015). This set of information is established at different moments over time, for example, for each medical consultation, laboratory test, and hospital admission. Nevertheless, there is a clear distinction between what was reported by health professionals and what the patient reports. Thus, the PHR offers an exact distinction between what was reported from each profile in its repository. In the case of a relative profile, some authors distinguish these profiles in terms of accessing the PHR with some limitations or full access with the permission of the patient (CASTILLO; MARTÍNEZ-GARCÍA; PULIDO, 2010), (HORAN; BOTTS; BURKHARD, 2010), (SUJANSKY; KUNZ, 2015). Additionally, in the case of public or anonymous profiles, the health data can be accessed in a limited or shared way, in which the PHR has a public and social nature to help other patients (WILLIAMS, 2010).

4.1.6.6 SQ4: What Are the Interaction Types of a Patient With a PHR?

In the SQ4 research question, we were able to identify three types of patient interactions with the PHR. In the first type, according to the definition of the PHR in ISO 14292 (ISO, 2012), the patient manages and controls the health data directly. In the second case, the patient only acts in a supporting role as complementation of EHRs but does not have effective control. Finally, in the third type, the patient outsources the management of the health data to a responsible person.

4.1.6.7 SQ5: Which Are the Techniques or Methods Used to Input Information Into a PHR?

Regarding the SQ5 research question, we sought to define the main techniques to input data into the PHR. As a result, with the analysis of the selected articles presented in Table 9, we can identify the techniques and profiles of the actors who use them. In the data collaboration (T1) technique, different health professionals access the PHR aside from the patient. The patient remains the PHR owner, but health professionals collaborate on input records in an identifiable and controlled way. In the second case, patient reports (T2), patients alone are in charge of inputting their medical record data without any support. In the third form, adaptive platforms (T3), the reported data, and the data collected from the EHR are integrated with the PHR data. In this case, data obtained from different sources and contexts are combined. The purpose is to provide better management of the patient's condition. For instance, it would be possible to provide real-time access to sensitive patient information and ease communication among patients and providers. In the case of the anonymization (T4) technique, medical data can be integrated with a social network, where the patient can share his or her status anonymously and receive contributions from other users.

4.1.6.8 SQ6: What Are the Goals of a PHR?

In the SQ6 research question, we sought to identify the PHR use purposes. This research question is related to the specific question SQ3, which aims to identify the objectives of the user profiles when accessing the PHR. We have identified three objective types. In the first case, the user profile accesses the PHR to only verify the health data without manipulating them. One example here includes health professionals or administrators who have permission only to view the data. In the second case, the user profile has permission to manipulate the data. In this situation, it is important to highlight the need to identify and control the profile that has changed the data and which data have been changed. In the third case, the user profile only monitors the records. An example of this might be a case in which the PHR receives data from sensors (IoT) and can send alerts depending on a situation.

4.1.6.9 SQ7: What Are the Types or Models of Architecture of PHRs?

Finally, in the SQ7 research question, we identified the architectures related to PHRs. We divided them into two groups: types (AG1) and coverage areas (AG2). In the case of architecture models, some articles state that health data are still stored on paper in many places, and other institutions have evolved into the proposed hybrid architectures with the PHR distributed inside and outside the health care organizations. In the case of the possibilities of coverage areas, we identified types ranging from a stand-alone PHR on a single machine to PHRs that can be taken from one country to another following an open international standard.

4.2 Related Work on Two-layer Architecture Model

This section presents related work to our OmniPHR two-layer architecture proposal. With more than 2500 studies returned in the search, we eliminated those works that do not deal directly about computer architecture models. Finally, we selected the most relevant regarding our research topic, which we highlight the related work in Table 12.

Table 12 – Related Work - Architecture models.

Related Work	Year	Model [1]	Interoperability	Security
HDEHR	2012	DE,P2P	-	-
m-Health	2013	DE	CCR	-
uPHR	2013	DE	HL7, CCR, CEN	-
CF	2014	CS,DO	-	CIA, HIPAA
HealthVault	2014	CS	CCR, HL7	Authentication
healthTicket	2014	CS	HL7, CCR	CP-ABE
DEPR	2015	DC	OpenEMR [2]	-
My HealtheVet	2015	DE	HL7	Security policies
SNOW	2015	DC	*open*EHR	Privacy policies

[1] Architecture Models: CS = Client-server;
P2P = Peer-to-peer; DO = Distributed objects;
DC = Distributed components; DE = Distributed event-based;
[2] Open-source medical software compliance with standards such as HIPAA, HL7, ASC X12 and SNOMED-CT.

Source: Prepared by the author

The works were evaluated from the architecture models according to the classification of Coulouris (COULOURIS et al., 2011), as well as regarding the security and privacy mechanisms that they use and what standards the model supports or is compliance:

- HDEHR (Hierarchical Distributed EHR) model (XIA; SONG, 2012) aims to maintain the patient's data in the health organization and replicate at the same time to other hospitals in their region, ensuring fail tolerance, but the P2P distribution is a future proposal and topics such as security, privacy, or interoperability are not covered.

- m-Health (Ubiquitous healthcare services in cloud) model (HE; FAN; LI, 2013) proposes an event-based distribution architecture, with services for interoperability following the CCR standard, although not mentioning about security or privacy.
- uPHR (Ubiquitous PHR framework) model (KSIMON; SONAI MUTHU ANBANANTHEN; LEE, 2013) is a distributed event-based model and has interoperability with HL7, CCR, and CEN 13606 standard, but also do not comment about security or privacy.
- CF (Conceptual Framework) model (SAFAVI; SHUKUR, 2014) is a framework to a wearable health system with a distributed mechanism based on cloud server distribution of objects with support for security and privacy with CIA and HIPAA protocols, but do not focus on interoperability.
- HealthVault (SPIL; KLEIN, 2014) is a proprietary solution following the CCR and HL7 standards. This is a web-based PHR to maintain health and fitness records but consists of a client-server platform where all health data are stored in the company's servers.
- healthTicket (KYAZZE; WESSON; NAUDE, 2014) is a design and implement case for ubiquitous PHR. This model is proposed as an architecture for patients' access by mobile and health care providers by a web application, following CCR and HL7 standards. This is a client-server model that uses a security mechanism called CP-ABE (Cipher-text Policy Attribute Encryption Scheme) to ensure privacy.
- DEPR (Distributed Electronic Patient Records) model (KEMKAR; KALODE, 2015) is a distributed components proposal based on OpenEMR system, which is compliant with several standards but does not focus on security or privacy.
- My HealtheVet (KLEIN et al., 2015) is an online tool for sharing health information and has a distributed event-based model with security policies and interoperability with HL7.
- Finally, SNOW project (HAILEMICHAEL; MARCO-RUIZ; BELLIKA, 2015) is a decentralized medical data processing system. This model uses distributed objects and has privacy policies following the *open*EHR standard.

4.3 Related Work on Performance Analysis

The selected related work studies are listed in Table 13, which lists the model name and reference, year of publication, health data standards, used framework, and if the study meets only organizational (EHR) or personal (PHR) health records. Table 13 underscores the fact that few studies dealt with the implementation of Blockchain technology applied to health records. Moreover, even fewer articles presented results with systematic quantitative evaluations. We analyzed the studies returned from these searches and selected only those studies that demonstrated Blockchain implementations involving health records in actual databases. We discarded studies that only conducted simulated evaluations, as well as those that only dealt with surveys or proposed solutions, i.e., without implementations that processed real data. Although the related work we examined was not restricted by date, we found relevant publications only from the year 2015 onwards since Blockchain technologies have just recently been explored in the context of health care.

Table 13 – Related Work - Performance analysis proposal.

Model & Year [1]	Health Data Std. [2]	FW [3]	EHR	PHR	Results
Invisible Ink, 2015	-	E	-	✓	Built Certified Mail service as sensitive user-data management platform.
FairAccess, 2016	-	E	✓	✓	Established an initial implementation with IoT and local Blockchain.
Healthbitt, 2016	HL7/FHIR, ISO13606	-	✓	✓	Stores patient data in a distributed ledger allowing sharing with doctors.
HGD, 2016	-	-	✓	✓	Potential way to house and share health care data.
MyData, 2016	-	-	✓	✓	Provides useful information on business models and ecosystems.
CBTi, 2017	-	H	✓	✓	Data update and evaluation process worked normally.
D-CAM, 2017	-	-	-	-	Adds a modest overhead and can be scaled for large networks.
MedRec, 2017	HL7/FHIR	E	✓	✓	Describes the technical design and early-stage prototype.
MeDShare, 2017	-	-	-		Comparable to solutions for data sharing between cloud services.
Patientory, 2017	HL7/FHIR	E	✓	✓	Potential to eliminate friction and the costs of third-party intermediaries.
Ancile, 2018	HL7	E	✓	✓	Discusses interactions with patient’s needs, providers and third parties.
FHIRChain, 2018	HL7/FHIR	E	✓	✓	Demonstrates a case study of collaborative app for remote cancer care.

[1] Models in ascending order by year.
[2] Health data standards.
[3] Platforms used in the solution, where E: Ethereum and H: Hyperledger Fabric.

Source: Prepared by the author.

4.4 Related Work on Multi-Blockchain Proposal

We submitted this search string to eight search portals for scientific studies in the health area, recognized by the academic community, without restricting the period: CiteSeerX, Cochrane Library, Google Scholar, PLOS ONE, ProQuest, PubMed, ScienceDirect, and Scopus. Initially, we found many related studies, mainly due to the restriction for searching articles that deal exclusively with implementations that refer to Blockchain-Distributed architectures. However, by removing articles that, in a first selection, dealt only with systematic reviews, contained only bidders without results, or even did not include the health area, we reduced the scope to less than a hundred articles. After this phase, we selected only those articles that had concrete results that were closer to our proposals, so that we could compare the proposals and the results. Then we select the following recent studies on the topic, according to Table 14. In the first column are presented the names or authors of studies selected as closer to this proposal. These selected studies allowed us to discuss the characteristics of models regarding the architectures, as well as the results obtained for comparative purposes of the projects. The second column lists the health data standards used or cited by the studies for structuring the health records, including *open*EHR (YANG; HUANG; LI, 2019) or HL7 FHIR (SARIPALLE; RUNYAN; RUSSELL, 2019). The third column contains the frameworks that served as the basis for proposals, including Ethereum or Hyperledger. Finally, in the last columns, it appears that the study deals with EHR or PHR. The blanks indicate that the proposals do not mention standards or type of health data.

Table 14 – Related Work - Multi-Blockchain proposal.

Model & Year [1] **Results**	**HDS** [2]	**BF** [3]	**EHR**	**PHR**
Bismuth, 2017	-	E H	-	✓
Assist in transition towards thoughtful and responsible data usage.				
SDN, 2017	-	E	-	✓
Performance improved reducing induced delay and ability to detect real-time attacks.				
SingularityNET, 2017	-	E	-	-
Ability to interface with multiple Blockchains.				
IoB, 2018	-	E H	-	-
Discuss interledger techniques for enabling industry-scale Blockchain networks.				
Sharma et al, 2018	-	-	✓	-
Resolves hierarchical security requirements with less consumption of energy.				
BDKMA, 2019	-	-	✓	-
Multi-Blockchain improves performance and scalability as network size increases.				
Hawig et al, 2019	F	-	✓	-
Designs presented suitable performance in enabling the interoperability.				
HCB-SDPP, 2019	-	H	-	✓
Can protect customer privacy more effectively than traditional smart home system.				
ReviewChain, 2019	-	E	-	-
Proposed smart contracts and notaries allow interfacing two Blockchains.				

[1] Models in ascending order by year.
[2] HDS - Health Data Standard, where O: openEHR and F: HL7 FHIR.
[3] BF - Blockchain frameworks or platforms quoted in the solution,
where E: Ethereum and H: Hyperledger.

Source: Prepared by the author.

4.5 Related Work on Semantic Interoperability

For the Semantic Interoperability Proposal, the methodology follows the principle used in the scientific community for designing a model, with the evaluation of a prototype system (GROENEN et al., 2016). The objective of this evaluation is to meet the health record interoperability requirements proposed for PHR. Regarding the research type, the approach is quantitative because the analyzed health data are from existing patients, although with anonymized data. As for the nature of the research, the study involves applied research because it aims at practical applications in the day-to-day lives of patients and health care providers. Regarding the objectives, the research used a case study applied to the context of the proposed model. As research materials to support our proposal, we start by investigating how studies of the past decade deal with semantic interoperability in health data.

Table 15 summarizes relevant related works that fit the concept of semantic interoperability according to the HIMSS. A reference model is a reference standard that uses a clinical information model (CIM), which is a structural standard of health data (MORENO-CONDE et al., 2015). As examples of CIMs there are the formats of templates used by the Health Level 7 (HL7) reference model, and the archetypes used in the *open*EHR and CEN/ISO EN13606 standards (MORENO-CONDE et al., 2015; SOCEANU, 2016; MARCOS et al., 2015). As can be seen in Table 15, practically all studies mention the HL7 reference model and the use of templates (Tm) in semantic interoperability (LAHTEENMAKI; LEPPANEN; KAIJANRANTA, 2009; GOOSSEN; GOOSSEN-BAREMANS; VAN DER ZEL, 2010; MUÑOZ et al., 2011; MO et al., 2015; ESPOSITO; CASTIGLIONE; PALMIERI, 2016; HEART; BEN-ASSULI; SHABTAI, 2017; PELEG et al., 2017).

However, among these works, only three studies mention Fast Healthcare Interoperability Resources (FHIR) (ALTEROVITZ et al., 2015; MANDEL et al., 2016; PAIS; PARRY; HUANG, 2017). The FHIR platform specification aims to promote and achieve interoperability among health systems using the HL7 reference model (BENDER; SARTIPI, 2013). However, only seven studies mention the *open*EHR or CEN/ISO EN13606 standards, and few works mention archetypes (Ar) associated with semantic interoperability. With one work found (ALYAMI; SONG, 2016), Dublin core (Dc) consists of metadata that can be used to retrieve and organize the PHR. Another study (HU; ELKUS; KERSCHBERG, 2016) mentions software agents (Sa), which consist of agent-based systems designed to interact and interpret health data.

Analyzing the selected articles, we observed a point in common among all studies. All research seeks to support semantic interoperability using one or more ontologies. Further, most studies cite terminologies (Tr) or vocabularies in health records, such as the LOINC, SNOMED-CT, and ICD standards, linked to the proposed semantic interoperability. However, we observed that few studies implemented or evaluated models with real patient data. Several articles consist of surveys or reviews, being limited to the conceptual description and presenting few concrete numbers regarding the results obtained.

Table 15 – Related Work - Interoperability

Reference & Year **Results**	**Reference Model**[2]	**Semantic Interoperability** [1]				
		Ar	**Dc**	**Sa**	**Tm**	**Tr**
Lahteenmaki, 2009	H	-	-	-	✓	✓
Occupational health pilot, with content from several apps merged.						
Goossen, 2010	H, O	✓	-	-	✓	✓
Two types of analyses and six initiatives evaluated accordingly.						
Muñoz, 2011	H, O	✓	-	-	-	✓
Description of how to achieve sharing and interoperability of clinical data.						
Alterovitz, 2015	H[3]	-	-	-	✓	✓
Feasibility shown by development of three applications.						
Marcos, 2015	H, O	✓	-	-	✓	✓
Implemented in two clinical domains.						
Moreno-Conde, 2015	H, O	✓	-	-	✓	✓
Common patterns to develop clinical information models (CIMs).						
Mo, 2015	H	-	-	-	✓	✓
Proposed 10 desired characteristics for computable phenotype repr. model.						
Alyami, 2016	H	-	✓	-	✓	-
Survey PHR in six categories and framework proposed.						
Esposito, 2016	H	-	-	-	✓	-
Proposes a semantic approach based on ontologies.						
Hu, 2016	-	-	-	✓	-	-
Presents an architecture for PHR recommendation.						
Mandel, 2016	H[3], O	✓	-	-	✓	✓
Relates development experiences and discusses challenges.						
Heart, 2017	H, O	✓	-	-	✓	✓
Reviews and presents needs to integrate EMR, EHR, and PHR.						
Pais, 2017	H[3]	-	-	-	✓	✓
Developed a conceptual model of wellness data using HL7 FHIR.						
Peleg, 2017	H, O	✓	-	-	✓	✓
Subsidized diagnosis change of 2/10 patients and anticipated therapy for 11/20.						

[1] Ar = Archetypes, Dc = Dublin core metadata, Sa = Software agents;
Tm = Templates; Tr = Terminologies.
[2] Reference Model (Standards): H = HL7, O = *open*EHR/CEN ISO EN13606;
[3] HL7 Fast Healthcare Interoperability Resources (FHIR) platform specification assists the solution.

Source: Prepared by the author.

4.6 Research Opportunities

We verified some research opportunities analyzing all the related work considered and presented in this Chapter 4. We identified several barriers to PHR adoption verifying the state-of-the-art regarding PHR through the systematic review. However, prominent challenges are related to the barriers regarding usability, portability, and performance of the solution. In addition, regarding the architecture proposal, we identified opportunities for research related to the confidentiality and privacy of health data, since it is a concern of patients, physicians, and IT professionals.

Regarding the interoperability proposal, another research opportunity detected is related to the inference of possible health problems from standardized health data. To close the scope of our proposal, this work aims to focus on two research opportunities found in the related works and that stand out: (a) problems related to the distribution of patient health data, where we propose a differentiated model of distributed architecture; (b) problems related to the diversity of existing health data standards, for which we propose a differentiated model of semantic interoperability.

As can be seen in the related work presented, there are several open research issues. Regarding the problems of architecture and interoperability, we can observe that there are several solutions proposals. However, we note that the use of Blockchain to address health issues of technology is a promising area where we can highlight the originality of our proposal. Our scientific contribution aims to promote the use of this type of technology as a model that addresses the problems related to the distribution of health records. We also see a number of open issues about the interoperability of health records, especially regarding semantic interoperability. This is an area with research potential, in which we also seek to highlight a scientific contribution, distinguishing our proposal for syntactic and semantic interoperability in this study.

5 OMNIPHR MODEL

This chapter details all parts that make up the OmniPHR model, with the functionalities and technologies that support the proposal defined previously. We start with an overview of the model, and then we present the OmniPHR model in detail. Regarding the model, we published two full scientific articles:

1. (ROEHRS; COSTA; ROSA RIGHI, 2017) in the Journal of Biomedical Informatics (JBI) on specifically the architecture model.
2. (ROEHRS et al., 2018) in the IEEE Journal of Biomedical and Health Informatics (J-BHI) on specifically the semantic interoperability model.

This chapter is divided into five sections, which are described below:

(i) Section 5.1 presents an overview of the OmnipHR model and the objectives.

(ii) Section 5.2 describes in detail the architecture of the model.

(iii) Section 5.3 presents the two-layered proposal for the OmnipHR model, specifying the use of Blockchain technology.

(iv) Section 5.4 describes an architecture proposal of the OmniPHR model, using a Multi-Blockchain setup.

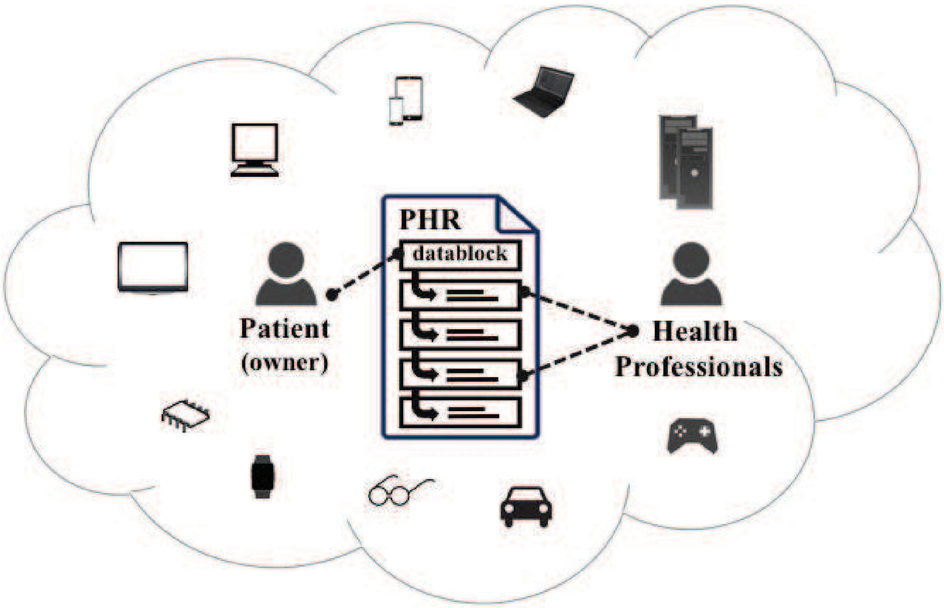
(v) Section 5.5 describes the OmniPHR interoperability proposal.

5.1 Overview

One of the conceptual bases of the OmniPHR is to divide the patient's health records into datablocks, which are a logical division of the patient's health datasets, such as laboratory, drug-related, x-ray, and others datasets, as can be seen an overview in Figure 6. As OmniPHR proposal is to promote an interoperable and distributed architecture for PHR with safety features for sensitive data, we found in Blockchain technology (NAKAMOTO, 2008) appropriate alternatives to composing our base architecture model. Furthermore, the OmniPHR model includes a premise of paging records, where the page's goal is that the user always accesses the latest data, but in a paginated way. This factor brings benefits ranging from faster access to the capacity of data maintenance be able, e.g., to keep older data in a less used repository.

OmniPHR focuses on the distribution and interoperability of PHR data. The model's purpose is to allow a unified view of health records, which are distributed in several health organizations, as well as address the challenges of having a distributed architecture that is scalable, elastic and interoperable. OmniPHR proposes a PHR representation, organized hierarchically, encrypted and distributed in chained datablocks on the network, using for this implementation

Figure 6 – Blockchain overview.

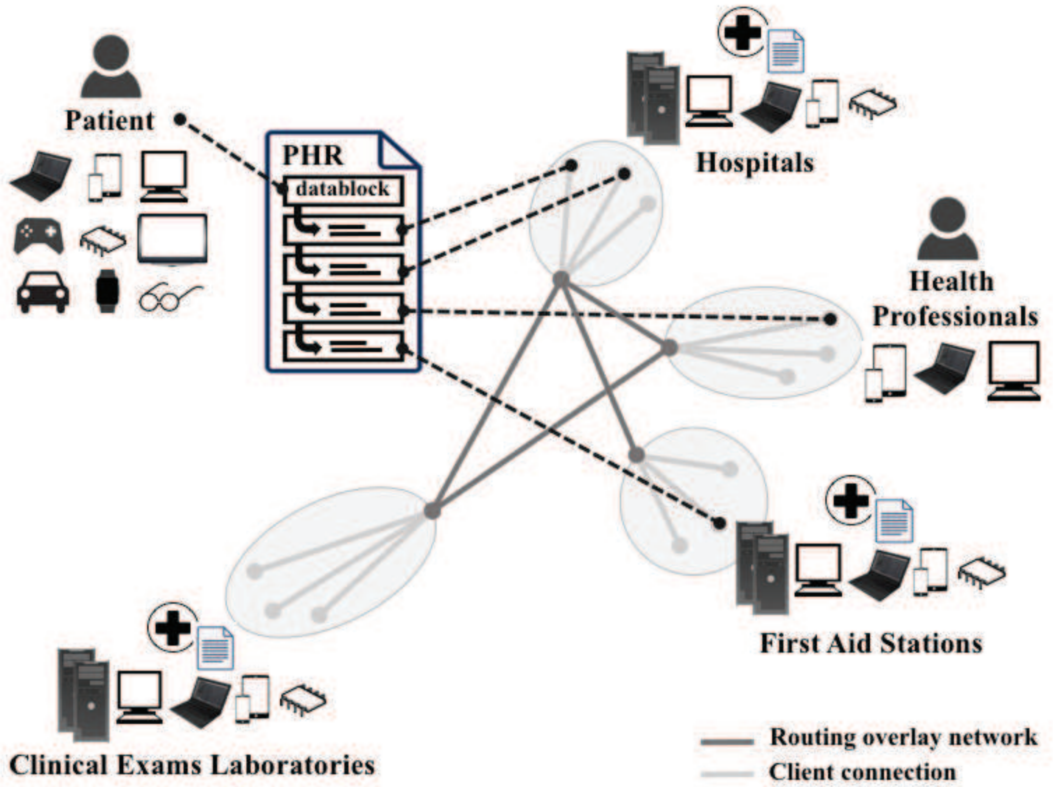


Source: Prepared by the author.

the Blockchain technology. These blocks can be located in different health care organizations and even in a patient-managed repository. In addition, the model provides the possibility of access by heterogeneous devices. In Figure 7, we observe a model overview. The figure shows a partitioned PHR in datablocks, distributed in a network with twelve nodes grouped into four subnetworks. It is possible to observe the diversity of devices able to interact with OmniPHR. The devices can join the system as providers (a) or consumers (b) since one of the model's premises is have OmniPHR present everywhere, patients may be, for instance at home, at work or in a hospital. As providers (a) users can use different devices that can supply data to compose the PHR. As consumers (b) users can use devices that can read the PHR data.

In OmniPHR we propose the use of a P2P network with routing overlay, which is an application server with defined responsibilities, including cloud computing features such as horizontal scalability and elasticity (ROSA RIGHI et al., 2016). Nevertheless, the main goal of OmniPHR routing overlay is to have the ability to maintain and locate datablocks of PHR when required and validate whether the chaining is intact or had some manipulation. Each OmniPHR datablock is encrypted and digitally signed by the responsible for inserting the information, which can be a health professional, patient, or whom the patient authorized to access their health records. This means that even patient's demographic data have one responsible, e.g., full name, birth date, gender, current address, or identification document numbers. Likewise, each diagnosis or laboratory test results datablock also have one responsible for this information with the digital signature respectively. In case of data coming from sensors, datablocks reported by these devices are also properly identified. Hence, the proposal is that any health datablock informed in OmniPHR is encrypted and has the informant with digital signature associated.

Figure 7 – An OmniPHR distributed in the network.



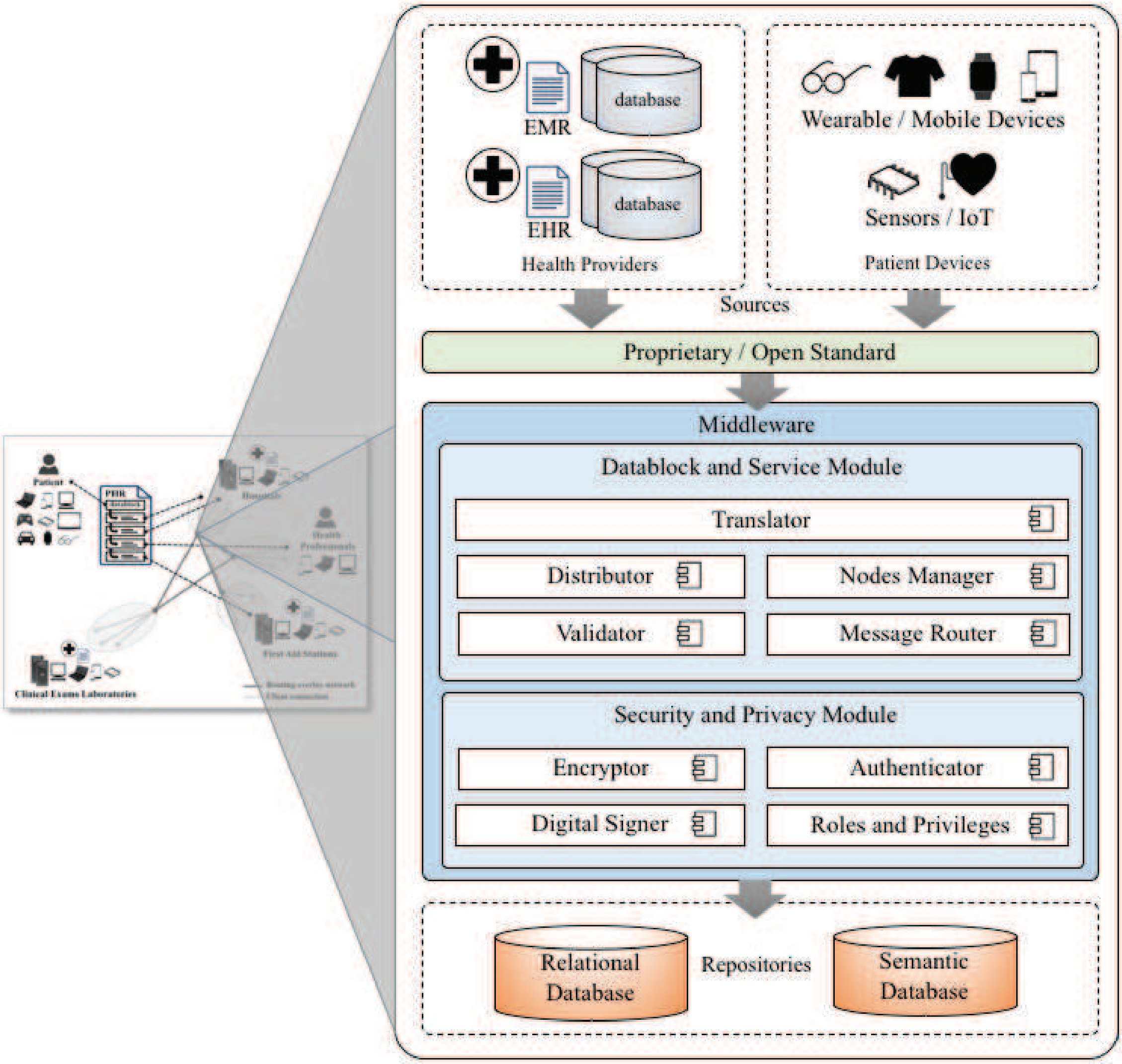
Source: Prepared by the author.

5.2 Architecture

In this section, we focus on the modules and components of OmniPHR design. The routing overlay node has a key role in the negotiation model design, acting as the main business component. The assignments are distributed in components split into three main modules, which are illustrated in the component diagram presented in Figure 8. In the diagram, we depicted the middleware present in each routing overlay, which is a logical abstraction of all modules and business components.

We can see in Figure 8, from top to bottom, that the data sources can be diverse. Health care providers may have a legacy database in a proprietary format (EMR) or following an open data standard (EHR). In addition, the data source may come from data collected on devices connected to the patient, such as wearables, or personal devices, such as mobile devices. As can be seen in Figure 8, in green color, the entry health records in the model middleware may follow a proprietary or open pattern. Also, in Figure 8, in blue color, the middleware and its respective components are presented within the OmniPHR model. Each module and component part is described following. Moreover, at the bottom of Figure 8, are the data repositories used by the OmniPHR model, which can be a relational database or a Knowledge Base (KB), represented by a semantic database.

Figure 8 – OmniPHR architecture model.



Source: Prepared by the author.

5.2.1 Datablock and Service Module

The Datablock and Service module is in charge of (1) translate, (2) distribute and (3) validate each patient's health datablock, as well as (4) manage the nodes and (5) routing services of connection messages. The proposal is to separate this module in components for each of these responsibilities that deal with the datablocks distribution and network management services. Following, we summarize each component:

5.2.1.1 Translator

The Translator component performs a key role in the OmniPHR since it is the input and output gateway of the datablocks. This component may be considered primarily responsible for interoperability in the model. By default, OmniPHR adopts an open standard for storing the health datablocks on superpeer. Thus, this component is only used when the health care provider uses a different standard of OmniPHR. That is, if the provider uses the same OmniPHR format, then this component is not triggered. On the other hand, if the provider uses another standard, whether it is open standard or not, then this component is triggered to translate the datablocks when they pass through the superpeer. The proposal consists in converting the altered parts at the source to the standard format adopted by the OmniPHR model. In this way, the component could promote interoperability with different standards of health datablocks, i.e., between the input and output standards.

In this process of constructing the conversion logic, the archetypes of *open*EHR standard can be organized in templates, collaborating to adapt the source data to its format. The component has the ability to translate datablocks in two ways: (1) in case the provider uses an open standard different from that adopted in OmniPHR; or (2) in case the provider uses a proprietary standard. In case of OmniPHR model, the proposal is the use of an equivalent ontology for each datablock, stored in a semantic database (as can be seen at the bottom of Figure 8), and using NLP to assist in automating the conversion of legacy health records to the standard format adopted by the model.

The next section, which deals with the semantic interoperability of the OmniPHR model, contains more details regarding the use of this component to promote the conversion of heterogeneous health records.

5.2.1.2 Distributor

After going through the Translator component and the content being translated and converted, we have the Distributor component, which is in charge of distributing and replicating datablocks on the network. The component requires knowledge of datablocks location, as well as the ability to fetch datablocks in the appropriate node that contains the requested data and return to requester. By default, the datablocks are stored on the computer where it was created, and some copies are distributed on the routing overlay and on the network following the DHT algorithm adopted.

We define this component in order to maintain a balanced distribution of data in the network. With this, the component also helps to minimize the risk of data loss, as well as enable a quick search of the data. In this way, the original data reported by a health care provider remains stored in the health organization with copies of these datablocks distributed over the network. In case of data informed by the patient, these are stored in the routing overlay, as well as with copies distributed in the network.

5.2.1.3 Validator

To help ensure reliable data, we have the Validator component, which has the responsibility of validating the health datablocks chaining. The tasks are to check the integrity of datablocks, to check and ensure the consistency, as well as the correct sequencing of datablocks. When a datablock is inserted in the chain, each one is formed by a dataset with time it was created and the hash pointer of the previous datablock. This principle was recommended by Satoshi's model (NAKAMOTO, 2008). However, each new datablock must be authenticated before it can form the next datablock in the chain, which is one of the routing overlay responsibilities through this Validator component. In this way, this component has the responsibilities of verifying the content consistency of each block, as well as the chaining with the adjacent nodes.

5.2.1.4 Nodes Manager

This component is one of the service components of the Datablock and Service Module. This component manages and controls the input and output of regular (or leaf) and routing overlay nodes in the network, promoting scalability and load balancing capabilities. For the input and output of network nodes, this component follows the rules of the DHT algorithm adopted by the OmniPHR model.

When a node wishes to be inserted into the network, this component generates a new identifier for the node according to the DHT algorithm and notify other nodes that this node is accessible. In the same way, if any node should be removed, the responsibility of this component is to command the redistribution of the nodes in the network, according to the DHT algorithm. In addition, if necessary, this component should trigger the distributor component to redistribute the data blocks in the network.

5.2.1.5 Message Router

This component provides communication services, such as packaging and routing of messages, receiving and forwarding requests to other modules and components. This component works in conjunction with the Nodes Manager and Distributor components to access the nodes in the network. In this component, OmniPHR proposes the use of an open cache solution, which aims to achieve better performance.

This solution follows the same DHT algorithm to distribute replicas of datablocks, maintaining in memory for a limited time the newly requested datablocks. In addition to routing requests on the network, this component also has the responsibility of sending simple queries (pings) to discover the nodes that are online on the network. This function aims to keep an updated routing table and speed up network communication.

5.2.2 Security and Privacy Module

The Security and Privacy module has a number of tasks regarding privacy and security maintenance. The responsibilities ranging from the protection of stored and transmitted datablocks through (1) encryption and (2) digital signature, promoting the privacy and data integrity, as well as a component for (3) authentication, until the control of (4) roles and privileges granted to the profiles. Following, we summarize each component:

5.2.2.1 Encryptor Component

This is the first component of the Security and Privacy Module, which has the responsibility of establishes the transmitted and stored datablocks encryption. This component encrypts datablocks pointers, as well as the health datablocks contents. In the OmniPHR model, in addition to the chaining of the datablocks are encrypted, the content of the datablocks must also be protected, both in repositories and in when they are transmitting in the network.

The purpose of the OmniPHR model is to promote the security of the datablocks chaining and the confidentiality of the patient's health data. The solution base is an open public key encryption, which generates two cryptographic keys: one public and another private. The private key is secret, and the public key is distributed with the patient identifier;

5.2.2.2 Digital Signer Component

This component is responsible for the digital signature of datablocks on the transmission and storage on the network. Each user has a digital signature, which is used to assign each datablock informed, respectively. The purpose is to verify that a transmitted datablock is an unchanged copy of one produced by the signer (COULOURIS et al., 2011). In addition, this component intends to contribute to identifying each responsible for inserting or updating the health datablock through a digital signature and making it possible to obtain the authenticity of the data reported.

5.2.2.3 Authenticator Component

This component ensures authorized access and proper attribution profile, as well as preventing unauthorized access, blocking and providing lost access recovery mechanisms. The purpose is that all network connections must be authenticated in order to minimize problems with unauthorized user access. When entering on the network, the user must have an ID generated for health records identification. The ID creation follows the OpenID code, which is used to identify users (OPENID, 2017), in order to avoid duplication of users (MCCOY et al., 2013). This ID forms the main health records identifier.

5.2.2.4 Roles and Privileges Component

This component is in charge of registration, concession, and maintenance of network access profiles. The algorithm model of the proposal follows the RBAC (Role-based Access Control) principle used by the scientific academic community (LU et al., 2015) (TSENG et al., 2017). This component has two approaches. The first approach is a personal purpose with an individual control of permissions granted to other users' access their PHRs. In this case, the patient may grant access privileges to their health records to health professionals or third-parties, as well as revoke at any time. The second approach is an organizational purpose where it is possible to create and maintain health professional profiles. The proposal is that each health organization should define the profiles and privileges of their health professionals. However, the master controller of PHR remains with the patient, following the first approach.

5.3 Two-layer Model

This section describes the OmniPHR architecture in relation to the two-layer model, where we apply Blockchain technology (ROEHRS; COSTA; ROSA RIGHI, 2017; ROEHRS et al., 2018). The model follows a distributed P2P network architecture with superpeers (COULOURIS et al., 2012). We expand and introduce improvements in OmniPHR's Blockchain-based architecture and implementation. In particular, the model deals with aspects focused on OmniPHR's Blockchain architecture and the impacts arising from the replication of health data.

OmniPHR's Blockchain architecture model is comprised of the following two layers:

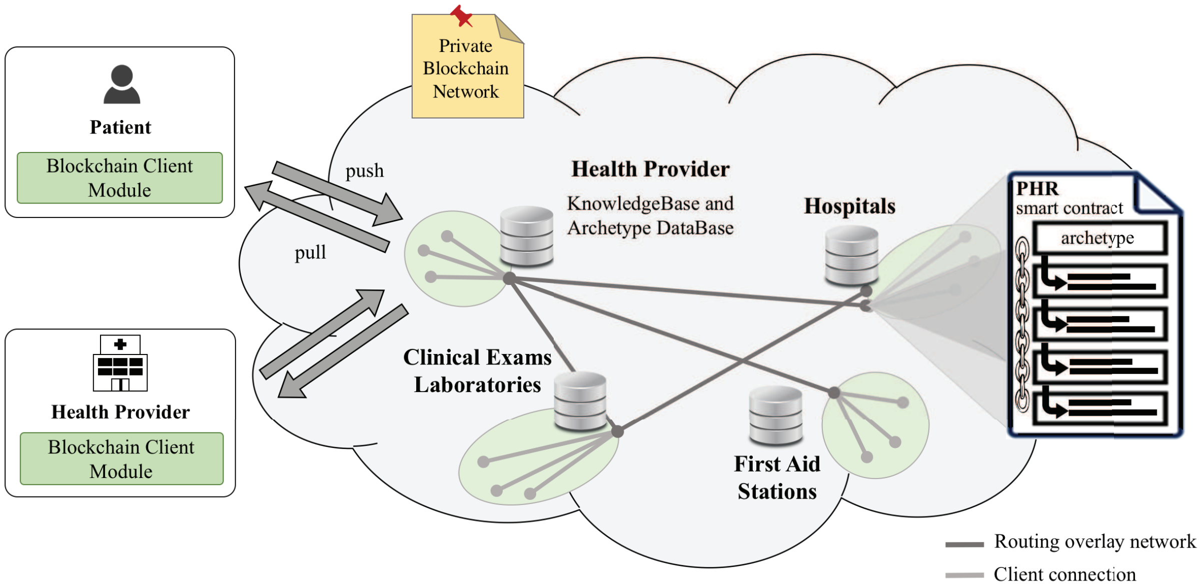
(a) **Client modules**, which are installed in the health providers and in the patient devices;

(b) **Server layer**, which is distributed in superpeers on a platform based on Blockchain.

This architecture is formed via a private P2P network, where health records are organized into datablocks comprising a linked list and a distributed ledger of health data (WALPORT, 2016). Figure 9 depicts the architecture of our OmniPHR prototype. This figure shows how clients communicate with the underlying Blockchain platform via pull and push messaging (COULOURIS et al., 2012). This format enables all clients connected in the network to update their data proactively, i.e., datablocks can be sent and received automatically. On the server, the Blockchain platform is installed on a set of distributed superpeers. This private network stores datablocks within a KnowledgeBase, which is a non-relational NoSQL database based on a Graph or RDF DBMS. The KnowledgeBase itself is implemented using the *open*EHR ontology to store the data in a non-relational database based on graphs.

Our OmniPHR prototype also uses a parallel database in an entity-relationship (ER) model to store the datablocks in the format of archetypes, which is a relational DBMS. These archetypes follow the *open*EHR health data standard, which we adopt for communication and data storage in our Blockchain network. The compositions of archetypes are the units that comprise the

Figure 9 – The architecture of OmniPHR prototype.



Source: Prepared by the author.

*open*EHR medical record structure (LI et al., 2018). The chained health datablocks in this database are used in forming the PHR smart contract.

Figure 10 shows how OmniPHR prototype chains health datablocks together. Each datablock consists of (a) content formed by an archetype containing the health record, (b) a field containing the hash code representing the digital signature of the content of the archetype, and (c) a pointer with hash code that set the previous datablock. The first datablock is named the 'genesis block', and the 'previous hash' field points to no other datablock since it is the first node in the linked list.

The OmniPHR prototype applies the Blockchain smart contract feature (SZABO, 1996) to verify and prevent violations of PHR data. Another highlight of our OmniPHR prototype involves the role of each node in the Blockchain network of health records. In particular, our prototype only allows superpeers located in the private network to evaluate the correctness of datablocks. Therefore, client nodes only consume microservices provided by superpeers. Moreover, clients also produce content that is evaluated and distributed on the Blockchain by superpeers.

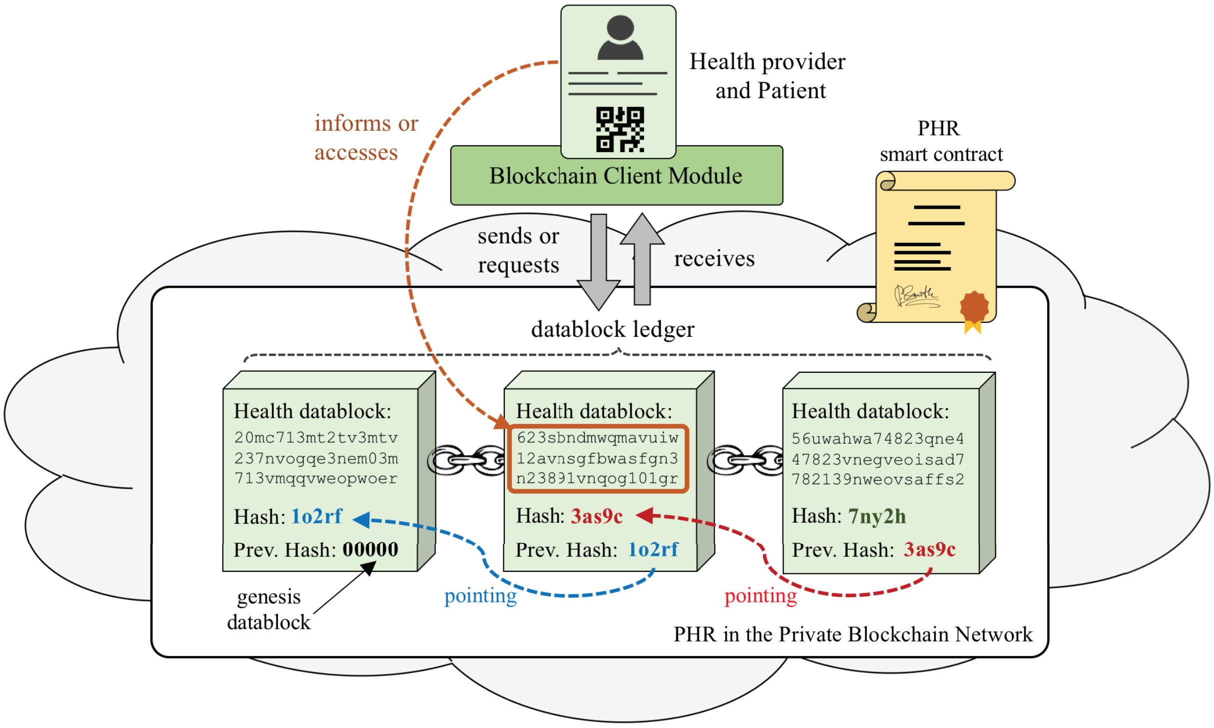
Datablocks in our OmniPHR prototype can be stored in the following two ways:

(a) Replicated in all nodes, following the approach adopted by the crypto-currency Bitcoin (ASPLUND; LOVHALL; NADJM-TEHRANI, 2018) or

(b) Using a replication algorithm, such as Chord (ROEHRS; COSTA; ROSA RIGHI, 2017), to replicate records only on certain nodes in the private Blockchain network.

The OmniPHR model can be configured to support both forms of replication because when using the Chord algorithm we can set up to how many nodes we want to replicate the data

Figure 10 – PHR Blockchain in OmniPHR



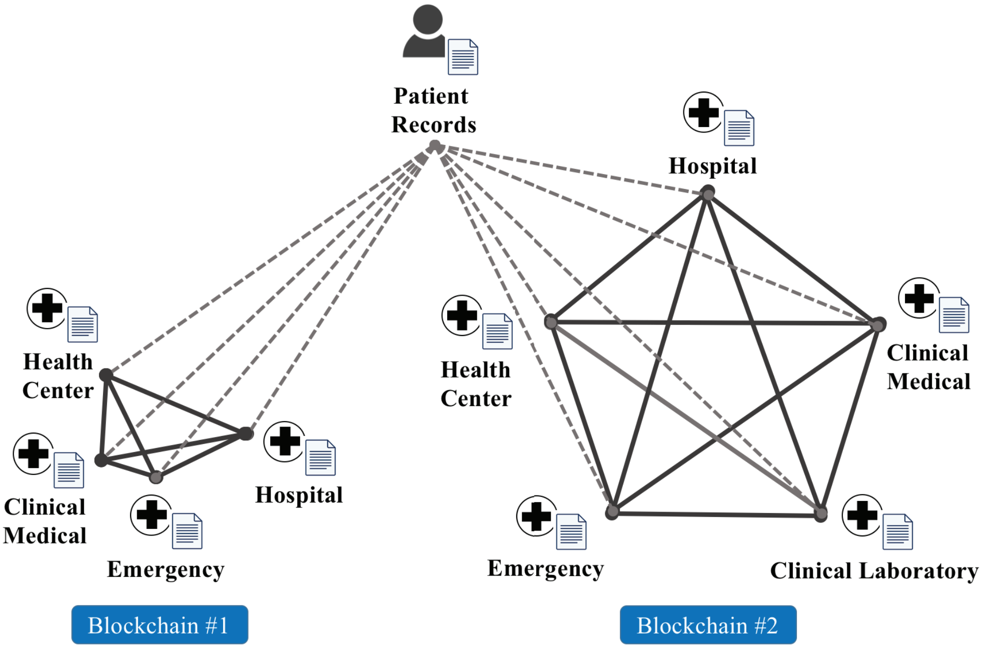
Source: Prepared by the author.

blocks. The Chord algorithm was used to make this decision flexible. This flexibility is one of the main characteristics of the model, since it may not be desirable or even performative to replicate health blocks for all nodes in the network.

5.4 Multi-Blockchain Model

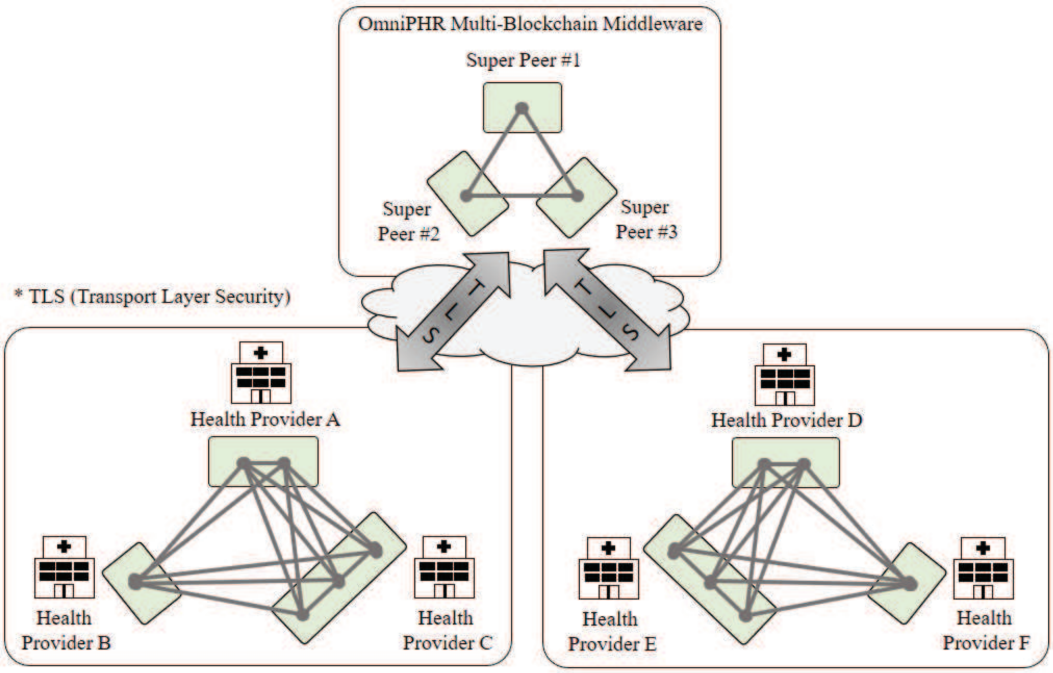
Initially, in the first studies about OmniPHR model (ROEHRS; COSTA; ROSA RIGHI, 2017; ROEHRS et al., 2018, 2019), we propose a model that used the Chord algorithm to perform the data replication for a limited number of nodes in a single Blockchain, instead of replicating to all nodes, as would the traditional setup of the Blockchain technology. As an extension, we propose a disruptive architecture, denominated OmniPHR Multi-Blockchain, that follows a different configuration compared to the traditional proposal of Blockchain. However, we remain committed to not replicating all data to all nodes and taking advantage of the distribution and security features of Blockchain technology, as can be seen in Figure 11. In this illustration, we can see that the PHR is distributed in different Blockchains (Blockchain #1 and #2), i.e., without replicating the data for all nodes, but maintaining the concept of integrated distribution. In the present proposal, we propose an architecture with multiple Blockchains orchestrated to support the distribution of PHR. Therefore, we use the traditional Blockchain architecture in a restricted context of nearby health providers, and a middleware orchestration to a context of distant health providers.

Figure 11 – PHR distributed in different Blockchains.



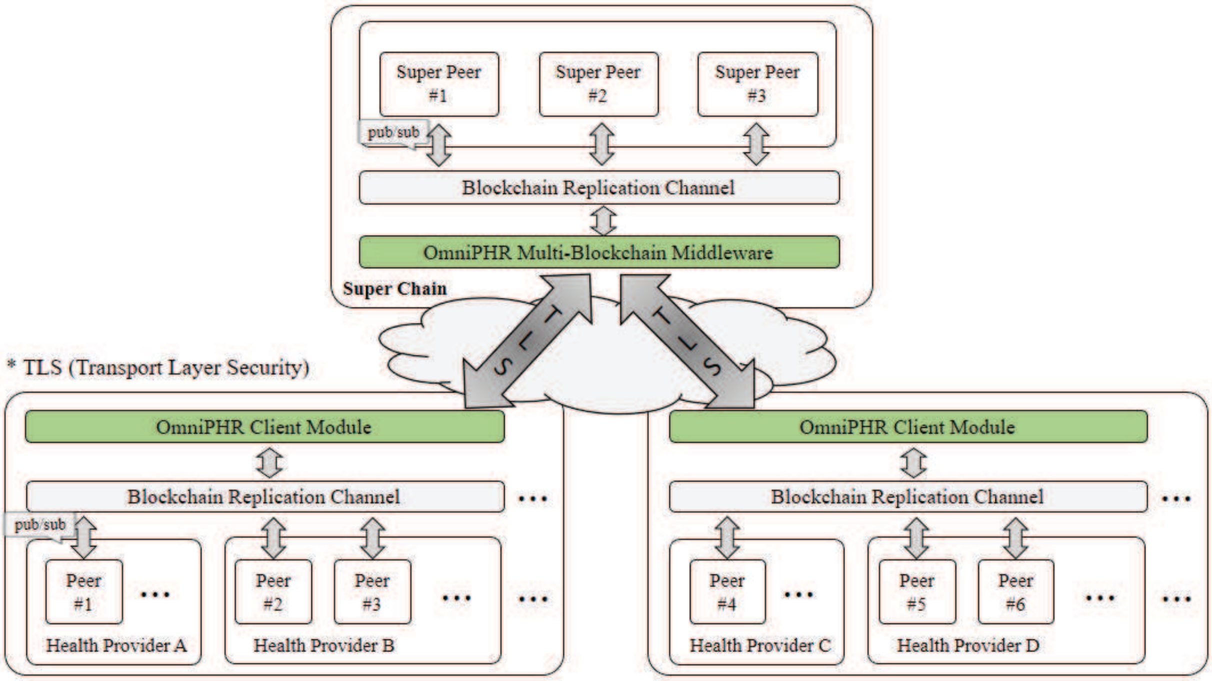
In this way, given the context initially presented, existing problems, proposed contributions, and related work, in this section, we detail our proposal of the architecture model. This proposal aims to support the implementation of Blockchain technology applied to health records in a distinct setup. Initially, we present a broad view of the architectural structure, as can be seen in Figure 11. Hence, we detail the pillars that form the technological and business view of the model.

Figure 12 – Overview of OmniPHR Multi-Blockchain architecture.



Therefore, given the distribution problem of health records of patients, we envision some aspects to address, which form the basis of our architectural model. In the Figures 12 and 13, we can visualize the base of the OmniPHR Multi-Blockchain model, understood by the questions concerning regarding (a) locality, (b) interoperability, (c) volume of data, and (d) security.

Figure 13 – Detailed view of OmniPHR Multi-Blockchain architecture.



(a) In terms of **locality**, this aspect concerns the physical location where the data is stored. A patient can knowingly go through various health providers throughout life. A patient can treat him/herself at home, in different places, in different countries, in different health organizations and by different professionals, as well as for short or long periods. Ideally, the PHR should collect the formal health records independently of location and timeless. This pillar is one where Blockchain's technology is best able to fit in, as it can promote an independent view of locality and at the same time linked health records through P2P networks. In this way, the Blockchain database of health records can promote a unified and distributed view of the data. This distributed format is a different model from the centralized one, where health records are concentrated in one location only and shared by health organizations. At this point, we note a distinction in our proposal regarding traditional Blockchain models, where all nodes share all data following the original model applied to the crypto-coins. Therefore, it is important to emphasize the aspect of data replication in the traditional Blockchain model, since data replication happens on all nodes of the network in order to keep everyone up to date (PARK et al., 2019).

(b) In terms of **interoperability**, this aspect concerns the standardization of the types and contents of the data, that is, it aims at meeting the great variability of types and contents that the patients' health records can store. Although the problem of data locale and its chaining we can suppress with the use of Blockchain, there are still other problems, starting with the nature of the data types in health records. The data entered patients' health records have several types. Several studies demonstrate the great variability of types of data that health records may contain. In essence, we can have structured and unstructured data records, which may even be in different languages, depending on where the patient was treated. In this sense, we have found the international interoperability standards HL7 (SARIPALLE; RUNYAN; RUSSELL, 2019) and *open*EHR (YANG; HUANG; LI, 2019), which can be integrated into some situations. Moreover, even based on the use of these international standards, there needs to be a concern with converting the data to the same protocol, since the variability of protocols around the world is wide, including private standards, as well as the variability of types of data is also vast.

(c) In terms of **volume**, this aspect addresses another fundamental aspect of the basis that composes a complete medical history of the patient and concerns the volume of these data. Many studies show that data records vary in volume, with media records, such as images and sounds, some of the most responsible for large volumes of data with regard to health records (KAUR; RANI, 2015). However, we can also have other types of records that demand large volumes of data like snippets of our DNA. In a structure that follows a centralized model this large volume of data may not be a big problem, however when it comes to a distributed model that follows a replication model for all nodes in the network, then this large volume of data can affect the system performance.

(d) In terms of **security**, this aspect concerns about data security in several respects. Some of the major concerns are about aspects such as access or privacy permissions; others relate to data breach or corruption, and other issues related to veracity or confirmation of responsibility identification for information. The use of Blockchain technology also seeks to collaborate on the security side, since the mechanism brings with it a model that, in addition to linking the data blocks, also aims to keep them inviolable. Hence, the concerns end up focusing on access and responsibility for the inserted data. In this sense, digital signature solutions can bring greater security in the composition of the identification of responsibility for the information inserted in the medical record.

In Figure 12, we present how to distribute the Blockchain network as an architectural model. We can see that there are two views on the distribution of records. Beginning with the internal view, this detailed view composes the internal Blockchain network of the health provider. This internal network aims to distribute the records in internal nodes of the health organization and in the nearby health providers, with all features that Blockchain technology provides in its original proposal, i.e., to replicate all data to all nodes of the network. This mechanism aims to facilitate the use of tools that follow the traditional Blockchain model, taking advantage of all P2P features and services.

On the other hand, as can be seen in Figure 13, we have the external view of the Blockchain network. This view aims to meet the access of a professional or even the patient himself/herself to data that are in other health providers. Therefore, this model aims to serve the data sharing between organizations and follows a different architectural model from the traditional Blockchain model, where all nodes receive all the data. Also on the external view, in this context, we seek to assemble the OmniPHR model using technologies that facilitate the integration and replication of data, such as the use of ESB (Enterprise Service Bus) with Publish and Subscribe, facilitating the interoperability of the model. This technological feature of the model aims to include both the attendance of possible different types of data coming from different organizations, as well as the integration with IoT devices and devices, for example.

The proposal format of the OmniPHR Multi-Blockchain model aims to take advantage of the original proposal of traditional Blockchain in internal use, with all its characteristics, thus facilitating the internal implementation by health providers. In addition, in the sense of the external use, the intention is to try to meet with a better performance the data replication, since a smaller number of replications of data we can realize, due to the size that a complete medical record can have over a lifetime.

5.5 Semantic Interoperability Model

This section presents the semantic interoperability vision of the OmniPHR, which provides to patients and health care providers the integration of different health data standards. Furthermore, OmniPHR also aims to aggregate semantic interoperability between the different PHR formats that the patient's health data can be structured. The architecture model can be illustrated according to the overview presented in Figure 14, with the OmniPHR middleware highlighted in green color, representing the core of the proposed model.

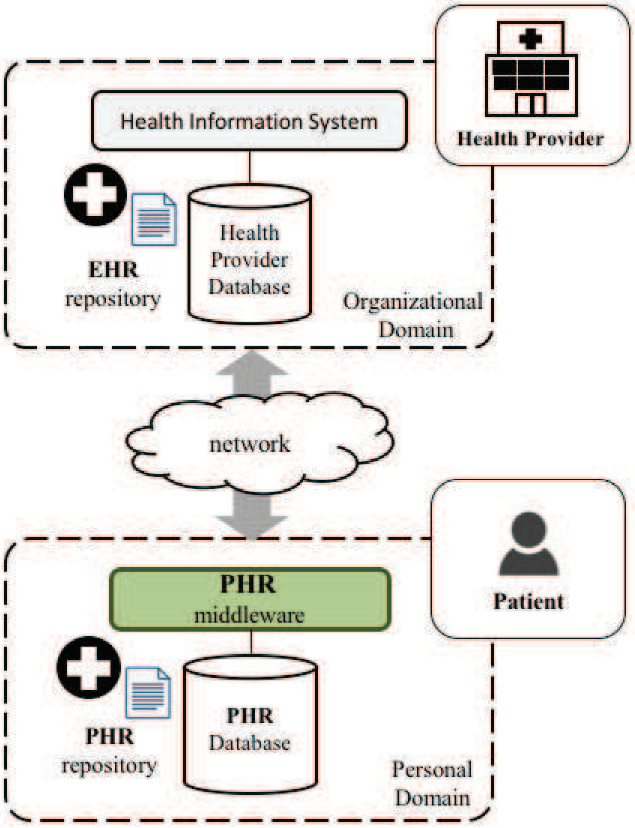
The model has two different domains. The first domain, the organizational domain, addresses the context of artifacts under the control of health care providers. The second domain, the personal domain, addresses the artifacts that hold the core of the model implementation. The organizational domain addresses the private context of health organizations. The proposal is to keep the original data contained in the databases of health care providers. Therefore, the health care provider can integrate its database with the model through two options:

(a) using a reference model that follows an open standard supported by the model, such as *open*EHR, CEN/ISO EN13606, or HL7 FHIR; and

(b) maintaining the current data definition standard, but providing the necessary subsidies for a semantic conversion of the internal model.

Figure 15 shows a detailed view of the model, following the division presented in Figure 14, with two domains:

Figure 14 – Semantic interoperability architecture overview.



Source: Prepared by the author.

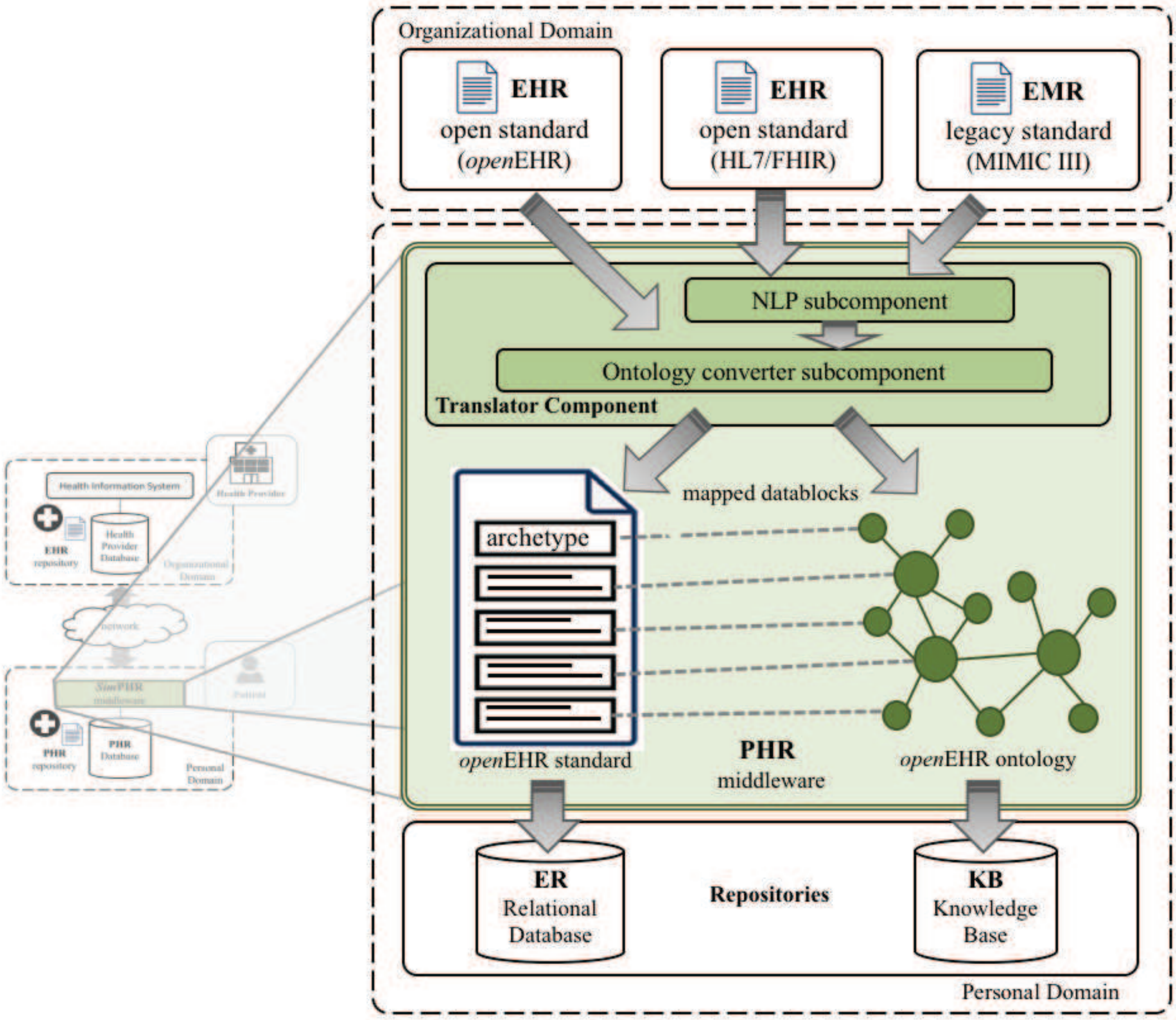
(a) one is the organizational domain, in which the objective is to preserve the health record structure used by the organization;

(b) the second is the personal domain, composed of the middleware, which is a business model layer and includes the repositories where PHR is stored.

In the organizational domain, OmniPHR predicts the entry of open and legacy standards. We have two representatives of open standards: (a) *open*EHR and (b) HL7 FHIR. To evaluate the legacy standards, we have one reference model: (c) Medical Information Mart for Intensive Care (MIMIC-III) (JOHNSON et al., 2016). In the middleware, there is a main component, the translator component, which has two subcomponents: (a) natural language processing (NLP) and (b) ontology converter. The health organization can submit to the OmniPHR middleware any of the three formats supported (*open*EHR, HL7 FHIR, or legacy), which read and convert to the *open*EHR ontology through the NLP processing phase. OmniPHR uses NLP resources to automate the conversion process. With ontology, we can integrate different standards, allowing the realization of inference about these data (OEMIG; BLOBEL, 2014).

After the health records are converted to the *open*EHR ontology, the data are stored in a semantic database repository, i.e., in a knowledge base (KB). In conjunction with this, the OmniPHR middleware replicates the health records, based on the *open*EHR archetypes, to the rela-

Figure 15 – Semantic interoperability detailed architecture.

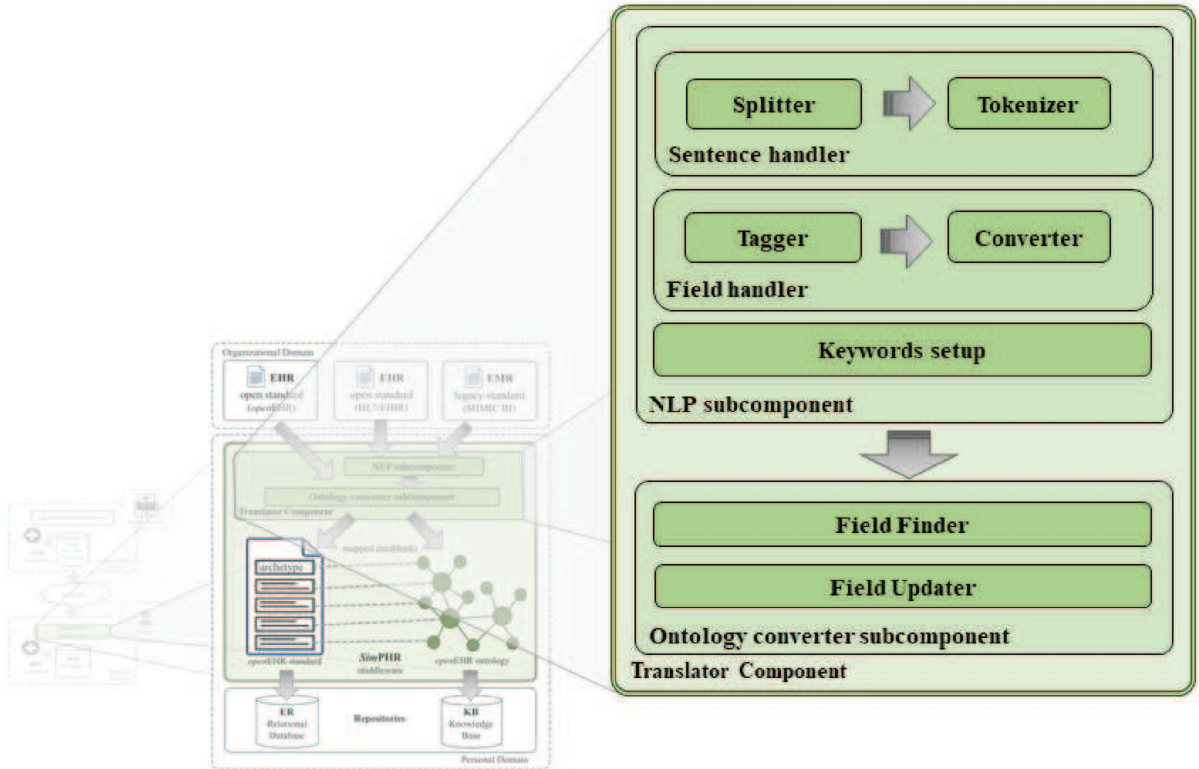


Source: Prepared by the author.

tional database. Considering the health records as controlled natural language (CNL), which is a language based on a certain natural language (KUHN, 2014), the main problem that OmniPHR addresses is extracting data from this CNL and converting it to *open*EHR ontology. In this way, the content that composes the PHR can be structured and unstructured data. The proposal is, besides promoting health record interoperability, to create the basis for enabling extraction and to infer possible health problems from the PHR unified viewpoint. We propose with OmniPHR a mechanism for health record conversion, using machine learning with NLP to automate the conversion to *open*EHR ontology. We have, in Figure 16, the details of the semantic interoperability model, where we can see the subcomponents of the Translator component. This is the main component responsible for the interoperability method.

The translator component is the main component responsible for the interoperability method. The method of converting the heterogeneous standards to the standard supported by OmniPHR begins receiving the corpus text and passing through the NLP subcomponent. The corpus can be a text in the XML or JSON format, represented by the HL7 FHIR, or CSV format repre-

Figure 16 – OmniPHR semantic interoperability method.



Source: Prepared by the author.

sented by the legacy MIMIC-III. According to the background of syntactic interoperability, we address the issues of syntactic interoperability through (a) functional and (b) data instance interoperability with the help of an XSD document (XML Schema Definition), which describes in detail the types and formats sent by sources. Considering the semantic interoperability issues with metadata, OmniPHR addresses the problem with training in the NLP component, through a neural network.

In the `NLP subcomponent`, the sequence of steps starts with the corpus text passing through the `Sentence handler`. This subcomponent has two stages, one that splits the sentence into words with `Splitter`, and the `Tokenizer`, which tokenizes the sentence identifying and separating relevant words from prepositions. Then, the text passes through the `Field handler` subcomponent, which has two functions: (a) tagging the words meaningfully through `Tagger`, identifying them according to the `Keywords setup`; (b) and afterward, the words are converted through the `Converter` to the standard used by the OmniPHR model.

At this point, the conversion to the *open*EHR archetype is complete. What remains is updating the OmniPHR model's ontology. The `Ontology converter subcomponent` is used to update the *open*EHR ontology, which has an object localization feature in the `Field handler`, and the feature used to update the ontology by adding, updating, or removing objects is called the `Field updater`. In this way, the steps of the working method of converting the heterogeneous standard to the OmniPHR standard are complete. OmniPHR updated the *open*EHR archetypes and the ontology, as well as stored them in the respective relational and KB repositories. In the following Figure 17 is the main algorithm that represents the extraction and conversion service in the Translator component:

Figure 17 – Algorithm for the extraction and conversion service.

```
Algorithm:  Extraction and conversion service.
input:  list of sentences (S) to translate
input:  setup of keywords (K) to verify
output:  ontology (O) filled
01 O ← queryCurrentOntology();
02 var:  current sentence (s) to convert
03 loop for each s ∈ S do
04     var:  current element (e)
05     e ← splitter(s);
06     e ← tokenizer(s);
07     loop for each e do
08         var:  current word (w)
09         w ← tagger(e, K);
10         w ← converter(e);
11         O ← fieldFinder(w);
12         O ← fieldUpdater(w);
13     end loop
14 end loop
```

Source: Prepared by the author.

The input parameters for the conversion algorithm are the list of sentences (*S*) to convert and the keyword setup (*K*). The output parameter and final goal of the algorithm is to fill the ontology (*O*). The execution begins by loading the current ontology using the `queryCurrentOntology` function and proceeds according to the steps of the translator component. In line 3 starts the main loop, which deals with the sentence, splitting the phrases and the tokens (words). Then, in line 7 another loop manages the conversion of words, finding the corresponding field to store the specific data. Finally, the ontology is updated, completing the process.

6 EVALUATION AND RESULTS

This chapter presents in sections the implementations and evaluations carried out on the Architecture, Multi-Blockchain, and Semantic Interoperability proposals, with their respective results obtained. We have published a full article (ROEHRS et al., 2019) specifically on performance analysis regarding the implementation of the OmniPHR model, based on Blockchain technology, in the Journal of Biomedical Informatics (JBI). This chapter is divided into four sections:

(i) Section 6.1 presents the implementation-related aspects of the tools used.

(ii) Section 6.2 contains the evaluation performed on the OmniPHR architecture proposal, and the subsections contain the detail of each step carried out in this evaluation.

(iii) Section 6.3 contains the tools and evaluation performed on the OmniPHR Multi-Blockchain proposal.

(iv) Section 6.4 contains the evaluation performed on the OmniPHR semantic interoperability proposal, as well as the subsections detail the steps taken to evaluate this proposal.

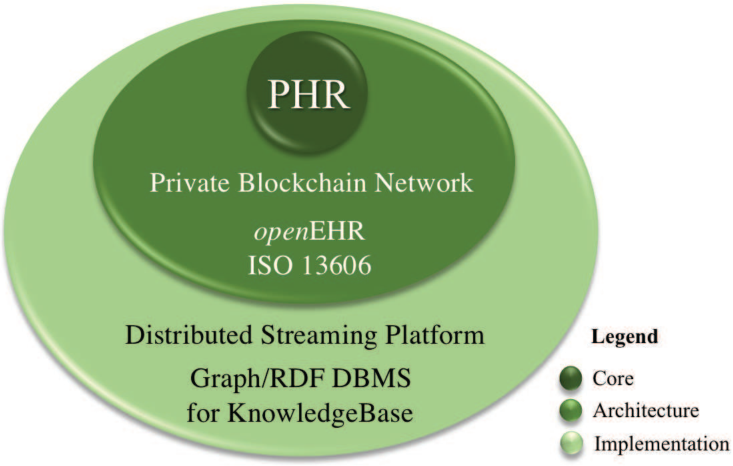
6.1 Implementation

A distinguishing characteristic of our OmniPHR prototype is its modular and distributed architecture based on components and microservices. We support the use of different components, as shown by the ecosystem in Figure 18. This Figure 18 should be viewed from the inside ring outwards. The core ring is PHR, which focuses on the integration of patient records. The second ring is based on a private Blockchain network and data protocol following the *open*EHR or ISO 13606 standard. The third ring used supports and implements the Blockchain network via a distributed streaming platform, as well as a graph-based database or RDF. This streaming platform enables the distribution and integration of health records, whereas the database in Graph or RDF format forms the KnowledgeBase ontology.

To support OmniPHR, we evaluated Blockchain platforms that have been applied to support health records, including Hyperledger Fabric (`www.hyperledger.org`) (ICHIKAWA; KASHIYAMA; UENO, 2017) and Ethereum (`www.ethereum.org`) (EKBLAW et al., 2016). To gain greater control, however, we developed our own Blockchain platform based on open APIs. This platform applies a private Blockchain format, i.e., a trusted network, where only clients who are authorized to participate can access health datablocks (DHILLON, 2017).

Table 16 summarizes all the platforms and tools employed in the OmniPHR prototype. We use the Apache Kafka platform to distribute the datablocks in the superpeers network (TA; LIU; NKABINDE, 2016). Kafka abstracts application concerns about data replication by extending its producer and consumer classes, which represent client nodes sending and receiving

Figure 18 – OmniPHR application ecosystem.



Source: Prepared by the author.

datablocks, respectively.

The Apache Kafka platform also acts as the message broker in the OmniPHR architecture, which uses its messaging and queuing features to exchange data between nodes. Its high-performance partitioning and replication capabilities are also used to support real-time processing systems. Apache Storm is a real-time distributed computing system associated with Apache Kafka. In contrast, Apache Spark supports large-scale data processing, making the OmniPHR architecture scalable and fault tolerant when distributing messages with health records.

Table 16 – Architectural choices.

Option	Potential benefits
Apache Kafka[1]	Distributed platform to store data safely in the distributed, replicated and fault-tolerant network.
Apache Zookeeper[2]	Configuration and synchronization services
Apache Storm[3]	Real-time computing for data stream distribution
Apache Spark[4]	Engine for large-scale data processing
OpenLink Virtuoso[5]	Multi-model DB, supporting KB and ER store

[1] Apache Kafka - https://kafka.apache.org/
[2] Apache Zookeeper - https://zookeeper.apache.org/
[3] Apache Storm - http://storm.apache.org/
[4] Apache Spark - https://spark.apache.org/
[5] OpenLink Virtuoso - http://sourceforge.net/projects/virtuoso/

Source: Prepared by the author.

We also use Apache Zookeeper in conjunction with the network resources provided by Apache Kafka. In particular, we use Zookeeper as an microservice interface to perform distributed configuration and synchronization of the messages that circulate in the Blockchain network (WANG et al., 2015). Apache Storm and Apache Spark services (JAIN, 2017; ZAHARIA et al., 2016) are also applied to support scalable and responsive processing needs. Our

OmniPHR prototype contains classes that serve as an interface to access the Blockchain, as well as store and remove content from the ledger. These classes enable the creation and maintenance of the PHR smart contract. Health data is stored in the open-source edition OpenLink Virtuoso database, which can store both relational storage (archetypes) and triple store (ontology) (ODGERS; DUMONTIER, 2015). The Virtuoso database enables data querying via the SQL or SPARQL (RDF) query languages. The OmniPHR prototype applies the Docker platform (`www.docker.com`) as the network container to provide a layer that abstracted and automated the virtualization (ADUFU; CHOI; KIM, 2015). To automate the building and deploying of code, we use Gradle (`gradle.org`) (IKKINK, 2015). To verify the transactions that circulate in the platform and to check with the content transmitted in the prototype, we exposed some microservices through RESTful web services, and we used the HTTP client SoapUI (`www.soapui.org`) to test the unification of health records. Finally, we used the Apache JMeter tool (`jmeter.apache.org`) to represent the concurrent load of client nodes by performing insertions of new datablocks in the network or queries of existing blocks on the network.

6.2 Two-layer Architecture Evaluation

Following the evaluation methodology proposed by Bossel (BOSSEL, 2013), which defines five steps to carry out the process we have to:

6.2.1 First - Developing the Model Concept

To illustrate the model behavior, we consider typical scenarios where patients go to various health organizations and are assisted by different health professionals. As a consequence, the patient's health records are updated many times. In this evaluation, we seek to assess the distribution and communication of health records in a network, following the proposal of OmniPHR model.

6.2.2 Second - Developing the Profiling Model

The objective is to evaluate the model initially in a setup with few nodes, which represent an initial situation of a health organization. Then, evaluate in some intermediate setups with varied settings, and finally, evaluate in a setup with a large number of nodes. At the end of execution, the purpose is to collect the averages of one-way hop count, messages present at runtime and one-way latency.

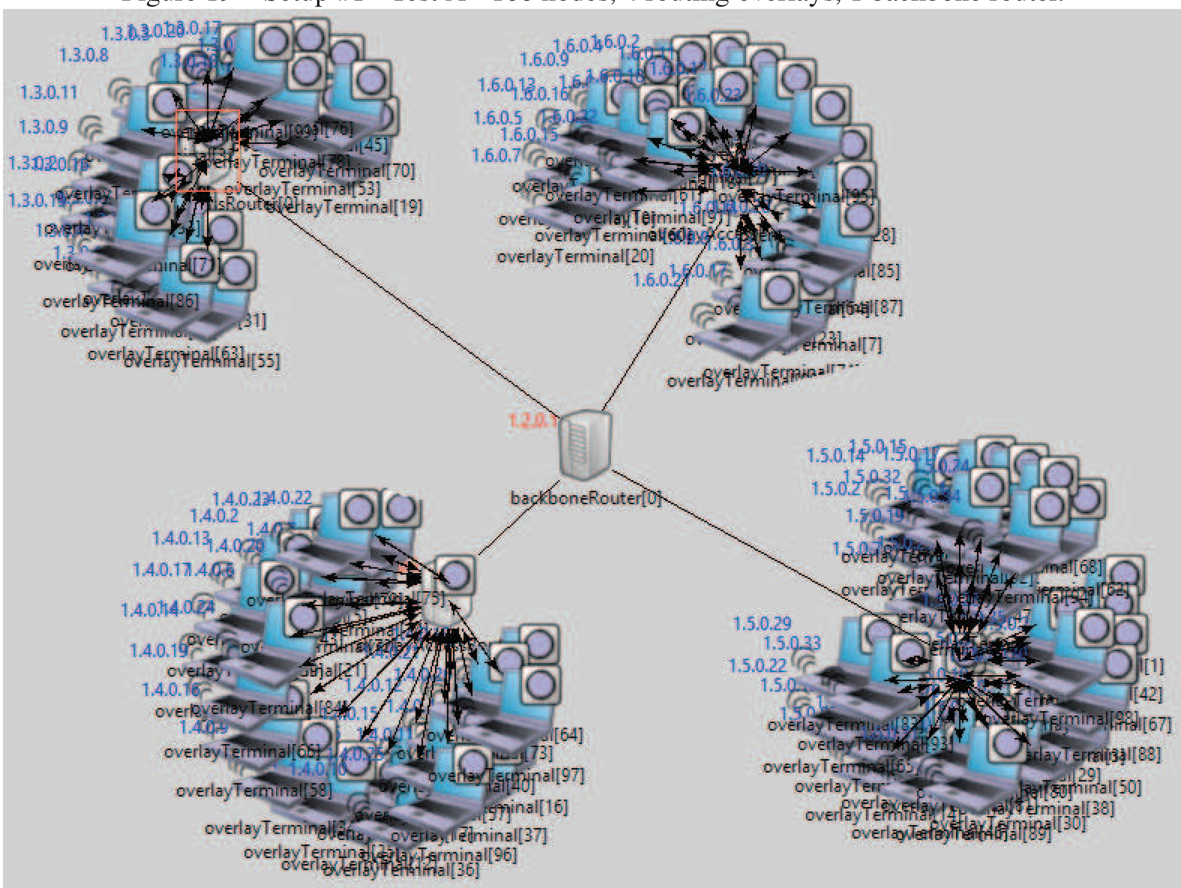
6.2.3 Third - Profiling of System Behavior

In this stage, we used as basis the OverSim framework (MOORHOUSE et al., 2013), which represents overlay and P2P networks. This framework is an implementation that uses the discrete event network environment OMNeT++ (JV; KALYANKAR; KHAMITKAR, 2014) and the INET Framework, which is an open-source suite of models for wired, wireless and mobile networks to OMNeT++.

6.2.4 Fourth - Performance Evaluation, Policy Choice and System Design

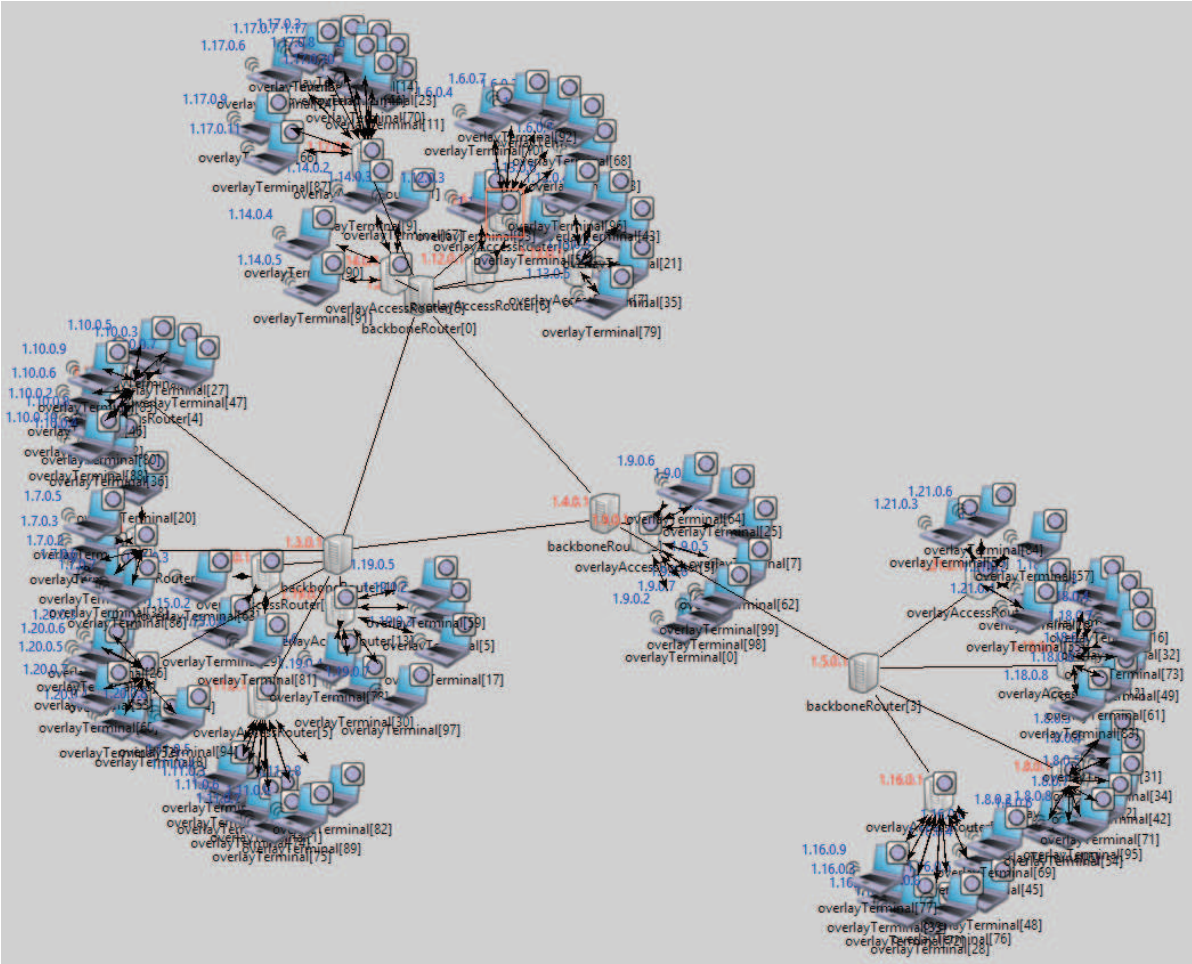
As environment settings, ten network setups with two different tests for each one have been executed, as can be seen in Table 17. To illustrate, a first setup (#1) in the ’A’ column with 100 nodes, 4 routing overlays, and 1 backbone router can be seen in Figure 19. The second test (’B’ column) with 100 nodes also, but quadrupling the number of routing overlays (4 to 16) and backbone routers (1 to 4) can be seen in Figure 20. The other tests followed the same logic, increasing the nodes, routing overlays, and backbones routers proportionally. In total, 20 tests were performed, i.e., 2 tests (A and B columns) for each setup of nodes.

Figure 19 – Setup #1 - Test A - 100 nodes, 4 routing overlays, 1 backbone router.



Source: Prepared by the author.

Figure 20 – Setup #2 - Test B - 100 nodes, 16 routing overlays, 4 backbone routers



Source: Prepared by the author.

All evaluations had a total period of 3 hours of execution and performed the following steps: (a) Entry of the number target of nodes for each execution in the network, with tests calibrated to have at most 5% above or below of entrances and outputs of nodes during the test period; (b) Random trigger of messages (each one representing one health datablock transmitted) at ranges up to 1 second concurrently between nodes.

6.2.5 Fifth - Mathematical Systems Analysis

In Table 17, we describe the parameters for each setup and the results obtained. In all evaluations, the use of CPU (maximum speed of 2 GHz with 2 cores) and memory (up to 8GB of RAM memory) was at most 50%. In the 'Setup' column, we have listed 10 test configurations. The 'Nodes' column refers to the target number of nodes in each evaluation setup, from 100 to 3200 nodes. In the set of columns 'Parameters' we listed the combinations of parameters tested, which are divided into the number of 'Routing Overlays' and the number of 'Backbones Routers'. The 'Routing Overlays' column refers to the fixed number of routing overlays nodes

Table 17 – Evaluation setups and results.

Setup	Nodes [1]	Parameters				Results					
		RO [2]		BR [3]		MP [4]		OHC [5]		OC [6]	
		A	B	A	B	A	B	A	B	A	B
1	100	4	16	1	4	1490	1570	4.17	4.60	0.216	0.260
2	200	5	20	2	8	2881	3138	5.03	5.02	0.247	0.393
3	400	6	24	3	12	6518	6593	5.32	5.34	0.479	0.403
4	800	8	32	4	16	13269	13457	5.54	5.57	0.491	0.475
5	1200	10	40	5	20	14561	15542	5.84	5.87	0.381	0.488
6	1600	12	48	6	24	19739	20091	6.03	6.07	0.564	0.445
7	2000	14	56	7	28	24975	25579	6.21	6.22	0.480	0.425
8	2400	16	64	8	32	30393	30802	6.34	6.35	0.549	0.547
9	2800	18	72	9	36	34511	36304	6.45	6.45	0.552	0.523
10	3200	20	80	10	40	40635	42035	6.52	6.56	0.470	0.555

[1] N = Number of nodes per setup;
[2] RO = Routing Overlays;
[3] BR = Backbone Routers;
[4] MP = Messages Present;
[5] OHC = One-way Hop Count;
[6] OL = One-way Latency.
Source: Prepared by the author.

in each setup and tests. The ’Backbone Routers’ column refers to the fixed number of backbones in each setup and tests. For each of the setups, we run two tests, according to the ’Test A’ and ’Test B’ columns. The ’Test A’ had the objective of verifying the network behavior with an increasing number of nodes per routing overlay, from 100/4 to 3200/160. For the second test, we quadruplicate the values of the ’Test A’ parameters. The objective of ’Test #2’ was to verify the network behavior with a smaller number of nodes per routing overlay, from 100/16 to 3200/80. In the set of columns ’Results’, we listed the results obtained for each of the setups and for each parameter used in ’Test A’ and ’Test B’. The ’Messages Present’ column refers to the average number of messages present on the network each moment, i.e., in transmission at every instant. The ’One-way Hop Count’ column refers to the average number of hops each message jumps between nodes in one way, i.e., between the source node and target node. Finally, the ’One-way Latency’ column refers to the average time of delay in seconds that a message took to traverse the network from the source node to the target node.

6.2.6 Environment for Evaluation Methodology

To help load the KnowledgeBase of health data, we used the CaboLabs EHRServer (GUTIéRREZ, 2018) platform. This platform implements the *open*EHR standard in a relational database. Using data stored in archetypes—and following the *open*EHR standard—we distributed the records into datablocks in the Blockchain.

To evaluate if the datablocks comprised a unified view of the health records, we evaluated the response time, the amount of memory occupied and the CPU usage, in a private Blockchain network with 10 superpeers and up to 40,000 concurrent sessions. That is since the used database has data of 40,000 patients, and as a way to perform a stress test on the system, we have tested the Blockchain to the limit of having at least one block of data from each patient searched or included concurrently. Each superpeer node consisted of Intel(R) Core(TM) i5, 3.30 GHz CPU, 4 cores, and 8GB RAM. We also profiled the OmniPHR prototype behavior by submitting different types of queries from an increasing series of client nodes.

Our evaluation environment used EHR and PHR for data query and health record manipulation (SHAH et al., 2016). As a load test scenario, therefore, we shared the use of the network Blockchain by having half the client nodes query blocks of registers and the other half insert blocks into the Blockchain network.

For comparison purposes, we created the following three test scenarios that performed an increasing number of queries and inserts operations:

(a) **Light scenario**, which had datablocks triggered from 50 up to 500 concurrent sessions in the network;

(b) **Medium scenario**, which had datablocks triggered from 1,000 up to 10,000 concurrent sessions;

(c) **Heavy scenario**, which had blocks of records transmitted from 13,000 up to 40,000 sessions on the network.

After configuring the settings to start each test scenario, we ran the network for nearly a week. During this period of $\sim$160 hours, we performed several load tests to evaluate the Light and Heavy scenarios. These load tests obtained the necessary values for the MTBF and MTTR calculations, obtaining results of 3.9586 and 0.0414, respectively. Based on these results, we calculated the Availability (A), where we obtain the value of 0.98964. The number of users accessing the network during the execution of the Light scenario was increased gradually, starting from 50 initial concurrent sessions until reaching the number of 500 users, as shown in Figure 21, which depicts the Light scenario results.

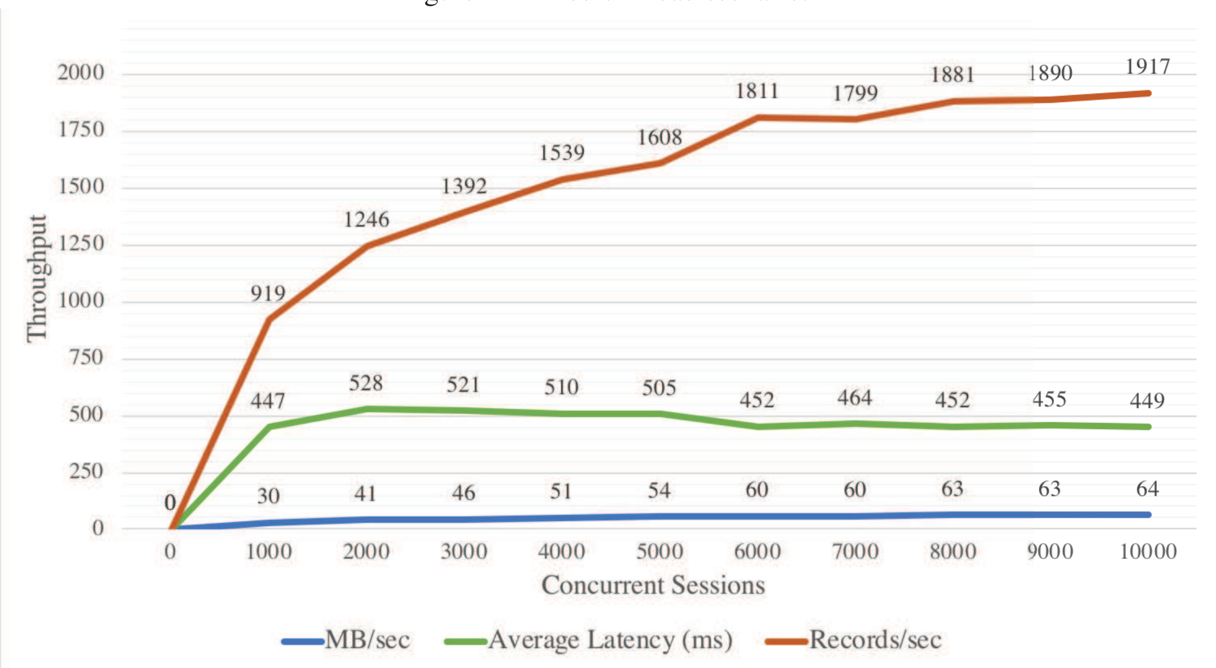
The average load of blocks transmitted in the Blockchain during the load test period is represented in Megabytes. The average response time (i.e., the average time of end-to-end latency that a client node requests to query a block or insert a new data in the Blockchain and obtain the response) is represented in milliseconds. Figure 21 shows the number of users accessing simultaneously the network in the Light scenario is increasing, as is the average load of records and the average response rate obtained. In this scenario, the load tests start from 50 concurrent sessions accessing the network, with a load of 5MB/sec of throughput, an average latency rate (end-to-end latency) of 52ms and 144 records processed per second, reaching 500 users (concurrent sessions), with 11MB/sec of throughput in the network, one average response rate of 214ms and 325 records/sec.

Figure 21 – Light load scenario.

Source: Prepared by the author.

In the second scenario of Figure 22, we can see a range from 1,000 to 10,000 concurrent sessions. Throughput ranges from 30MB/sec to 64 MB/sec. We can observe that latency is stable, almost unchanged, going from 447ms to 449ms, i.e., less than half a second. And the number of records per second goes from 919 to 1,917 records/sec.

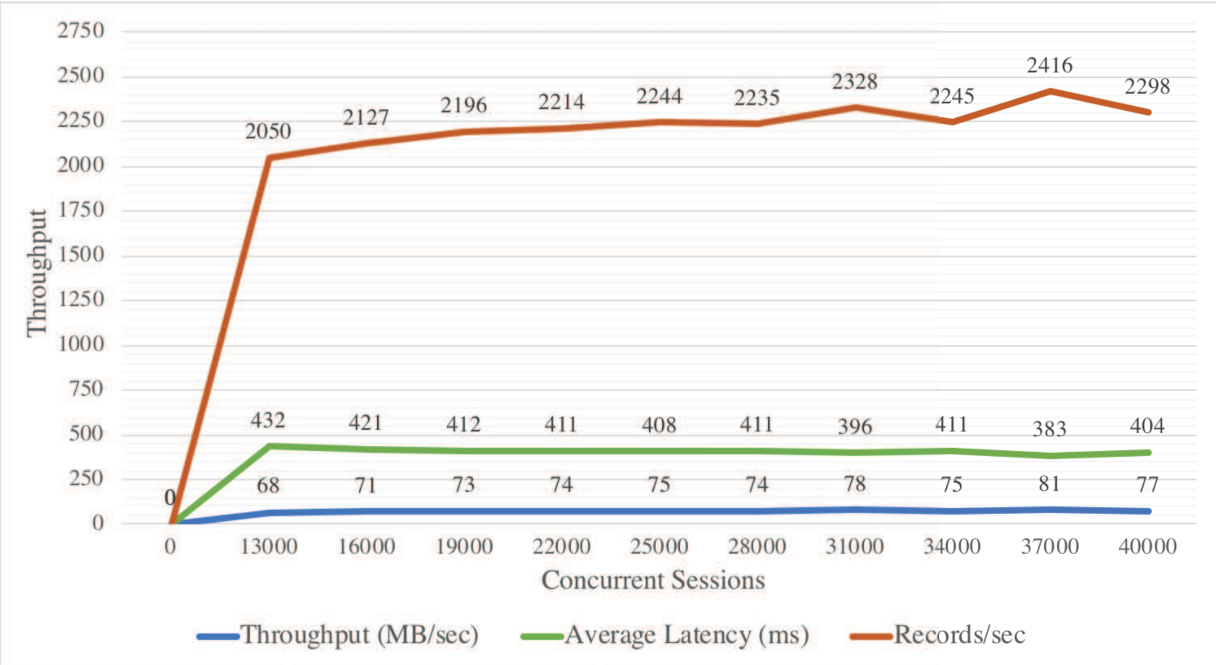
Figure 22 – Medium load scenario.



Source: Prepared by the author.

In the third scenario, represented in Figure 23, we can see the results from the Heavy scenario. This scenario also shows an increasing number of users, the average load of datablocks, and response rate. The initial load was 68MB/sec with the response time of 432ms for 13,000 concurrent sessions until 40,000 were reached, with 77MB/sec of throughput, an average response rate of 404ms and 2,298 records/sec. We can observe that even by increasing the number of concurrent sessions and throughput, the average response time remained stable.

Figure 23 – Heavy load scenario.



Source: Prepared by the author.

Table 18 presents data collected in the load test profiling for other non-functional requirements. The items analyzed were (a) CPU Usage, (b) Memory, (c) Disk throughput, (d) Network throughput (Sender) and (e) Network throughput (Receiver), for each of three scenarios evaluated (Light, Medium and Heavy). The variations of the data obtained in our tests for these requirements did not significantly impact the performance of the superpeers, except in the case of the heavy scenario, where there was greater use of machine resources.

Table 18 – Performance scenarios - average usage value per node.

Rated item	Light Load	Medium Load	Heavy Load
CPU usage average	0,3 GHz (10%)	0,75 GHz (25%)	1,5 GHz (45%)
Memory	0,8 GB (10%)	2,08 GB (26%)	3,6 GB (45%)
Disk throughput	0,1 MB/sec (0,1%)	4 MB/sec (4%)	10 MB/sec (10%)
Network throughput (Sender)	0,1 MB/sec (0,1%)	4 MB/sec (4%)	10 MB/sec (10%)
Network throughput (Receiver)	0,4 MB/sec (1,5%)	2 MB/sec (10%)	4,5 MB/sec (21%)

Source: Prepared by the author.

After we applied the methods for evaluation, the results from the MTBF and MTTR calculations comprised and demonstrated a 98% solution availability during load tests. These results were obtained by subjecting the model to three scenarios: one light with until 500 concurrent sessions accessing the network, one medium with up to 10,000 sessions and one heavy with up to 40,000 sessions. The scenarios used the same amount of patient data.

6.3 Multi-Blockchain Evaluation

Regarding the Multi-Blockchains-based model, we performed some experiments in order to verify the efficiency and performance of the proposal in a context with real data from a database of approximately 40 thousand patients. We perform the tests within a week of submissions of inserts and queries of records in the same Blockchain network and on another Blockchain through the OmniPHR Multi-Blockchain middleware. In the context of clients (for internal view), representing the health providers use, we tested with two Blockchains, each blockchain having two health organizations and each health organization with two peers. In the context of SuperChain, representing the OmniPHR middleware (for external view), we tested with two super-peers in the Blockchain. The configuration of each superpeer in the SuperChain context was an Intel(R) Core i5, 3.30GHz CPU, 4 cores with 8GB RAM. In the client context, each peer was an Intel(R) Core M, 1.1GHz CPU with 8GB RAM.

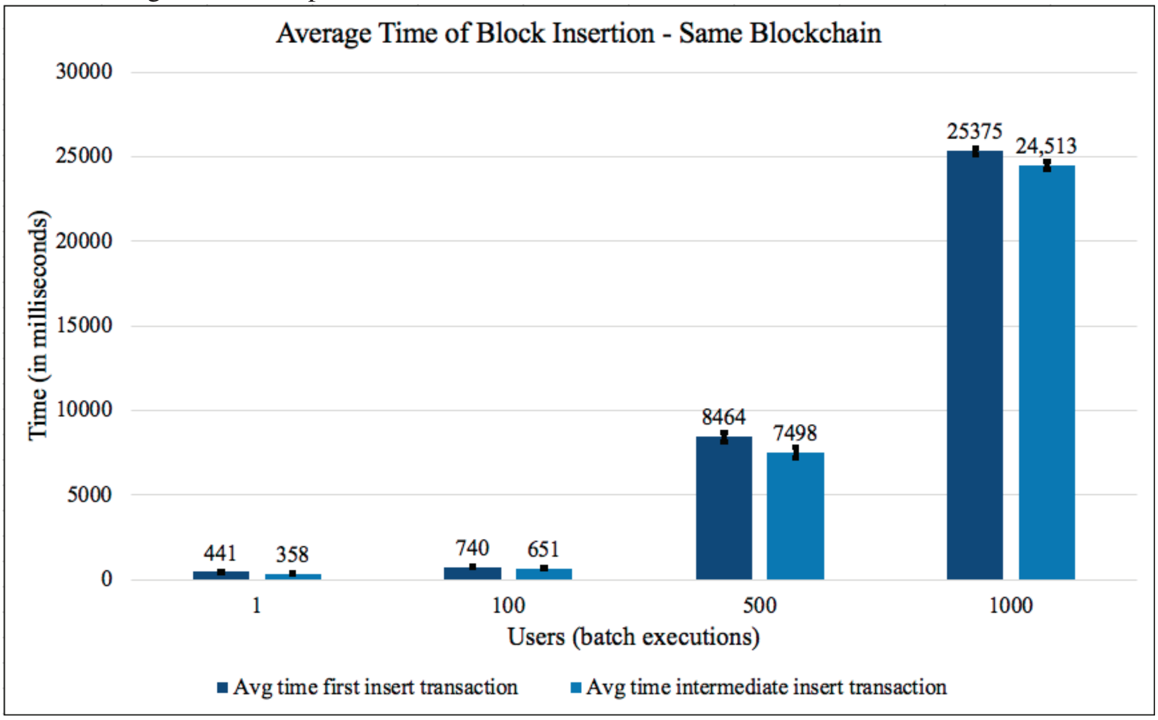
We perform the transactions of insert and query in datablocks, i.e., we work with parts of a PHR, following the model of archetypes of *open*EHR. We submitted the operations both in the internal view of the model, representing the use within a nearby network of health providers, i.e., same Blockchain, as well as in the external view, of data sharing between distant health providers, i.e., on another Blockchain.

In Figures 24, 25, 26, and 27, we present the graphs of results, with the response times, the respective standard deviation indicated and throughput. Moreover, on each graph, we present detailed numbers that compose the results.

In these results, we can observe that the first executions carried out in the same Blockchain and queries executed in another Blockchain with one hop. This means that in our tests, we perform the operations of inserting records only within the internal Blockchain, as well as queries. In addition, regarding other Blockchains, we only executed queries.

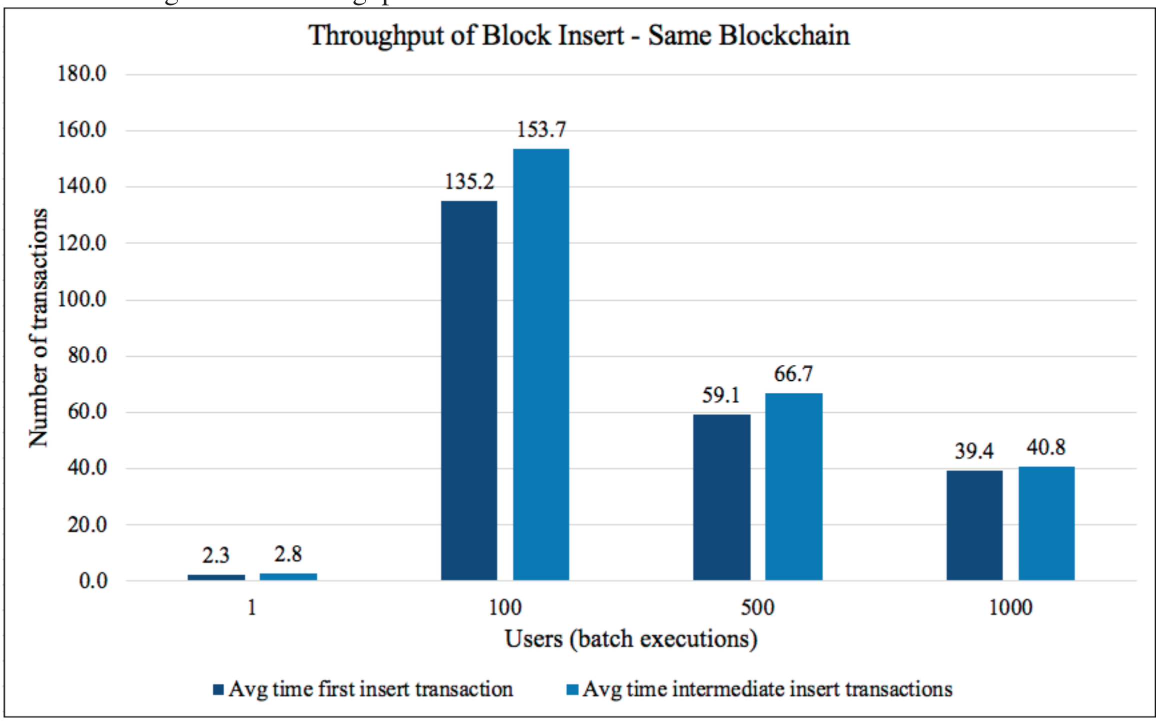
These test scenarios represent scenarios where health providers can only enter patient records within the internal Blockchains, never updating or removing record blocks. Similarly, health providers can only query records from other Blockchains, which means that health providers cannot enter or change records in other Blockchains. The principle is that healthcare professionals insert records in the Blockchains where the patient is close, and conducting queries on records on nearby and distant Blockchains networks.

Figure 24 – Response time of datablock insert transaction in local Blockchain.



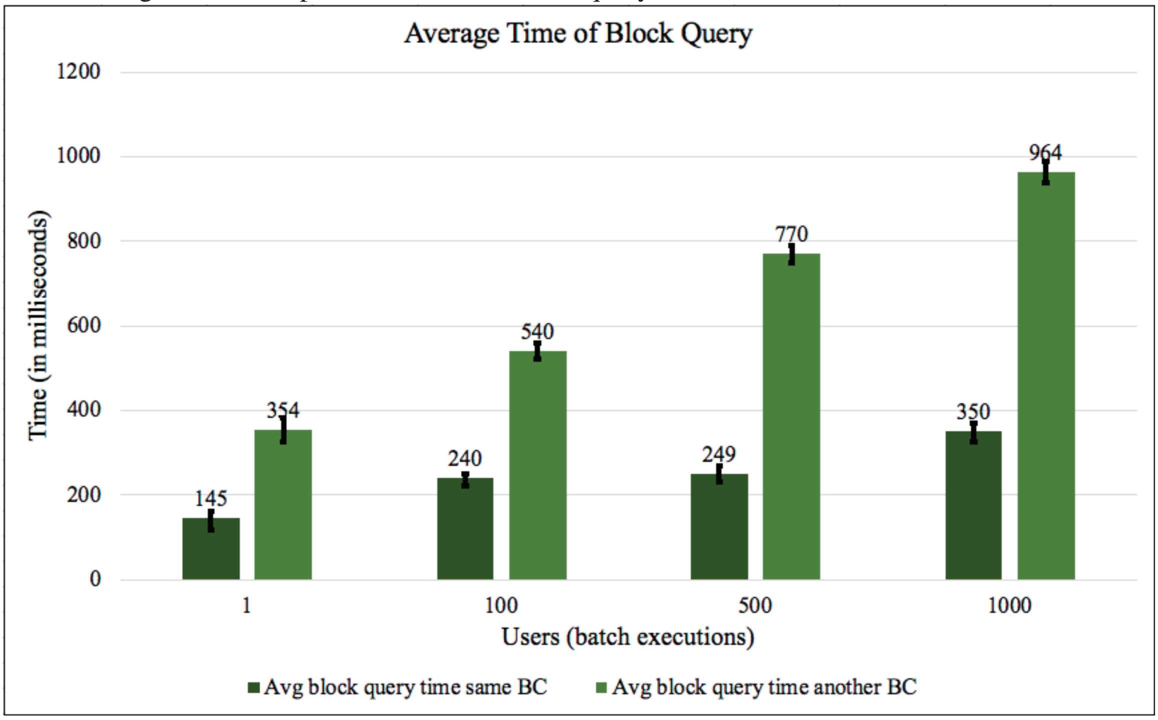
Source: Prepared by the author.

Figure 25 – Throughput of datablock insert transaction in local Blockchain.



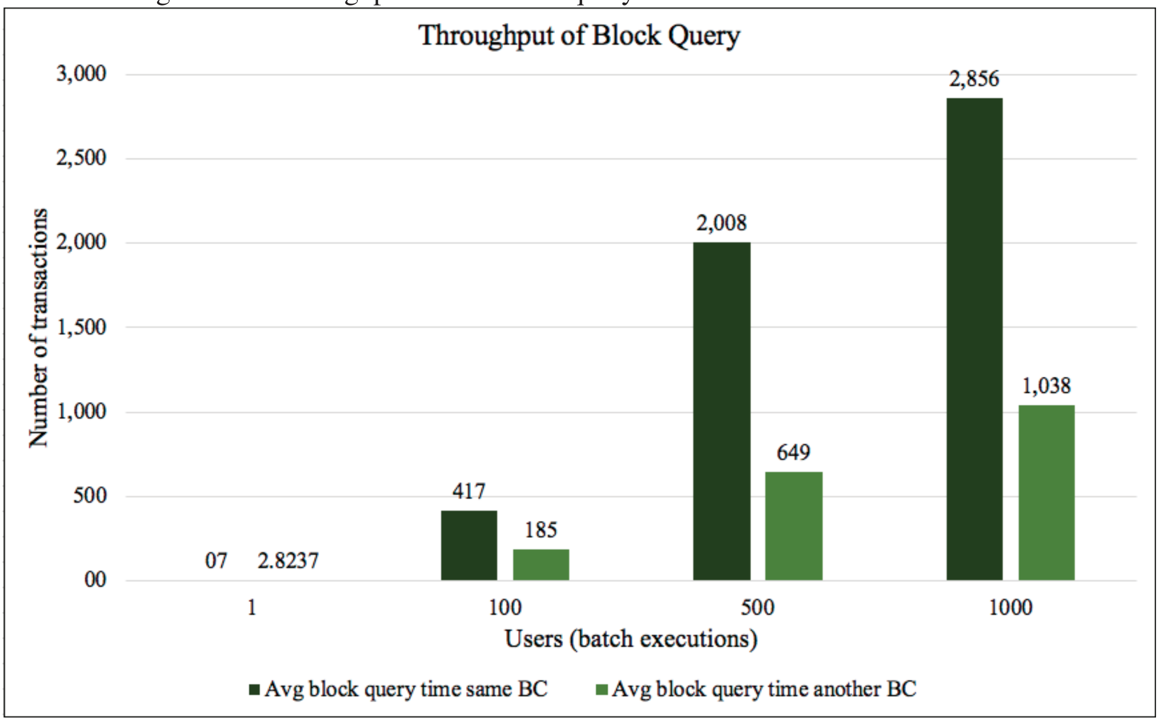
Source: Prepared by the author.

Figure 26 – Response time of datablock query in local and external Blockchain.



Source: Prepared by the author.

Figure 27 – Throughput of datablock query in local and external Blockchain.



Source: Prepared by the author.

6.4 Semantic Interoperability Evaluation

Analyzing the related work for the semantic interoperability proposal in Table 15, we could observe that, although some studies propose to integrate the heterogeneous standards using solutions such as ontologies, software agents or DC metadata, there are still gaps in the integration process. In addition to the diversity of existing standards, as the standards are generally incompatible (EDEN et al., 2016), one of the difficulties encountered is how to automate this process, since the standards and health records can change any time.

6.4.1 Selection of the reference model standard

Analyzing the two reference models for health records structure, *open*EHR and HL7/FHIR, we chose to use *open*EHR standard. This choice is due to the flexible combination of archetypes, which gives dynamism to the records structure. Another factor is due to the existence of ontologies for this standard (OPENEHR, 2017). The use of *open*EHR ontology aims to promote the use of a single open language that provides interoperability between heterogeneous standards. This also aims to provide the ability to infer and prevent possible health problems that the patient may have. Since the model uses the *open*EHR standard internally, if the health organization uses the same standard, then there is no need to use the NLP subcomponent. In order to automate the translation process of the heterogeneous health standards to the standard used by the model, the translator component uses NLP to perform the conversion to the *open*EHR ontology. This translation subcomponent is required when the health organization uses a different standard than the one adopted by the model, whether it is an open standard (such as HL7/FHIR) or legacy standard (represented in our model by the MIMIC III (JOHNSON et al., 2016)).

6.4.2 Use of Natural Processing Language and Ontology

In order to automate the conversion of the heterogeneous standards to the health standard adopted by the OmniPHR, we investigated NLP and CNL (Controlled Natural Language) parsing solutions that could be integrated with ontologies. Table 19 summarizes the tools found.

Table 19 – Natural Language Processing (NLP) parsing tools.

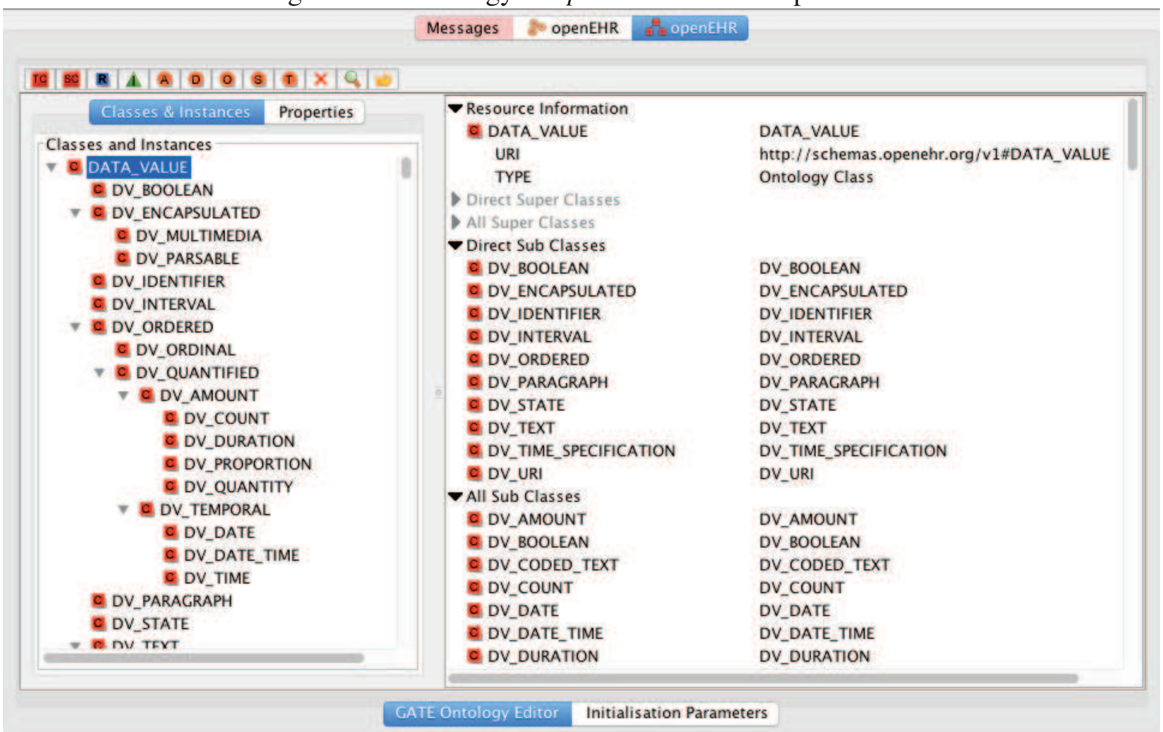
Tool and References	Date [1]	Citations [2]
Apache cTAKES – 4.0.0 (CTAKES, 2017) (SAVOVA et al., 2010)	2017-04-25	384
Attempto (ACE) – 6.7 (ATTEMPTO, 2017) (MENDES et al., 2014)	2013-10-04	1,130
GATE – 8.4.1 (GATE, 2017) (CUNNINGHAM et al., 2013)	2017-06-09	2,710
OWL API - 4.1.0 (OWL, 2017) (AZEVEDO et al., 2014)	2016-02-02	691
Stanford Parser - 3.7.0 (PARSER, 2017) (AZEVEDO et al., 2014)	2016-10-31	6,550

[1] Last stable release date; [2] Number of citations, searching by full tool name in Google Scholar.

Source: Prepared by the author.

We can see in Table 19 the name, references, date of last stable release and number of usage citations of the tool, where we can see a large number of uses. The Apache cTAKES (CTAKES, 2017) tool focuses on KB extraction from the Electronic Medical Record (EMR) analysis. The Attempto Controlled English (ACE) is focused on CNL, serving as a language for knowledge representation (ATTEMPTO, 2017). GATE (General Architecture for Text Engineering) (GATE, 2017) can be considered a platform, which provides several components and plugins for NLP and CNL parsing with manipulation of ontologies. The OWL API (OWL, 2017) is geared towards creating, manipulating, and serializing OWL ontologies. The Stanford Parser (PARSER, 2017) is dedicated to the parsing of structured, semi-structured, and unstructured texts. Due to the range of functionalities available for manipulation and processing of natural languages, mainly in relation to the populating ontologies, as well as due to the current state of the tool, which remains up-to-date, the proposal selected the GATE platform (GATE, 2017) (CUNNINGHAM et al., 2013). The GATE platform provides mechanisms that promote the interpretation and conversion of controlled text to an ontology, with a set of tools for NLP, such as: tokenizers, taggers and parsers (GATE, 2017) (DAVIS, 2013). Figure 28 presents the *open*EHR ontology imported on GATE platform (GATE, 2017).

Figure 28 – Ontology of *open*EHR on GATE platform.



Source: Prepared by the author.

6.4.3 Selecting the Data repository

In the OmniPHR architecture, we have PHR repositories divided into two instances:

- (a) one for the relational repository of the *open*EHR archetypes;
- (b) other for the semantic repository of the *open*EHR ontology.

To meet this demand, we evaluated four data repository solutions, according to Table 20.

Table 20 – Data repository solutions.

Database	Reference	Version	Relational	Graph	RDF [1]
Neo4j	(NEO4J, 2017)	3.1.3		✓	
PostgreSQL	(POSTGRESQL, 2017)	9.6.3	✓		
Stardog	(STARDOG, 2017)	4.2.4		✓	✓
Virtuoso	(VIRTUOSO, 2017)	7.2.4	✓	✓	✓

[1] Resource Description Framework.

Source: Prepared by the author.

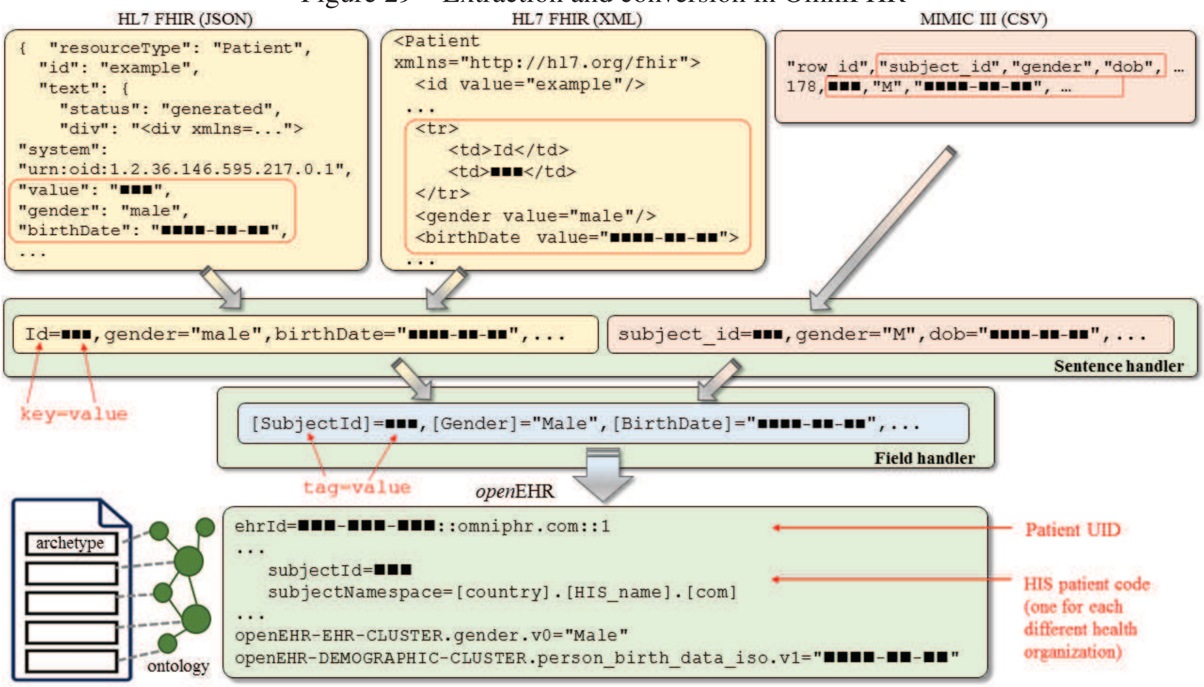
Neo4J (NEO4J, 2017) is dedicated to graphs and with the Cypher language is possible to make inferences about the graph data, although it does not have native support for RDF (Resource Description Framework) triple store, which is the basis for storing ontologies. PostgreSQL (POSTGRESQL, 2017) does not have native support for RDF, although it has support for XML, which is the RDF base format. Stardog (STARDOG, 2017) is dedicated to graph types and RDF triple store. Finally, OpenLink Virtuoso (VIRTUOSO, 2017) is a multi-model database, with relational data, graphs, and RDF triple store support. Due to the characteristics of providing access, both relational and RDF triple store, allowing the realization of inferences natively, we selected the OpenLink Virtuoso database version 7.2.4.2.

Analyzing the recognized reference models *open*EHR and HL7/FHIR, we chose to use the *open*EHR standard because of the flexible combination of archetypes and the existence of ontologies for this health standard. The use of *open*EHR ontology aims to promote the use of a single open language that provides interoperability between heterogeneous standards. The *open*EHR standard also integrates with vocabulary and medical terminology such as SNOMED-CT, LOINC, ICD, and ISO. In addition, this aims to provide the ability to infer and prevent possible health problems that the patient may have (ISO, 2014). Since the model uses the *open*EHR standard, if the health organization uses the same standard, then there is no need to use the NLP subcomponent. To automate the translation process of the heterogeneous health standards to the standard used by the model, the Translator component uses NLP to perform the conversion to the *open*EHR ontology. This translation subcomponent is required when the health organization uses a different standard than the one adopted by the model.

We used anonymous patient data available in the MIMIC-III database (JOHNSON et al., 2016) as input data to OmniPHR prototype. We used the version 1.4 of MIMIC-III, with 38,645 adults patients. We represented each patient with the standards supported by the model, i.e. (a) *open*EHR, (b) HL7 FHIR, and (c) legacy (MIMIC III). The MIMIC III database has real patient health data, although anonymized. In case of *open*EHR standard, we used the EHRServer platform through the EHRCommiter component to populate its database in PostgreSQL with data extracted from MIMIC III. We used this component to extract data in XML format. We obtained a document in the *open*EHR format generated from a solution that admittedly follows this standard. In case of HL7 FHIR standard, we followed the same script, using both the API and documentation in XML and JSON format, available on the FHIR website. In case of patient data that represented the legacy standard, we extracted data directly from the MIMIC-III database for plain text, in CSV format. OmniPHR receives, interprets, and tags all three formats through the NLP component. Then, OmniPHR converts the sentence to the *open*EHR archetypes and ontology model.

In Figure 29, we can observe that in the HL7 FHIR standard the patient identifier is `Id`, whereas in MIMIC-III it is `subject_id`. The *open*EHR standard allows mapping the patient's original identifiers in the Health Information System (HIS), referenced as `subjectID`, to a common identifier, called `ehrID`. The principle is that the identifier in *open*EHR follows a universal identification pattern (UID) (INFORMATICS, 2018). In this way, we can maintain the interoperability of patient identification codes.

Figure 29 – Extraction and conversion in OmniPHR



Source: Prepared by the author.

The first step is performed by the `Sentence handler` subcomponent, which reads the original format and parses to a key=value sentence. The second step is performed by the `Field handler` subcomponent, which identifies the fields and uses the tagger to create a tag=value sentence. Thus, the `NLP subcomponent` has the ability to learn, with a machine-learning algorithm, about input formats, and identify them the next time an equal input occurs. We evaluated the model using the Virtuoso database, which we populated through the OmniPHR. OmniPHR populates the relational portion of the database with the *open*EHR archetype data and the KB portion with the *open*EHR ontology data. We employed SQL queries to check the consistency of the replicated data in OmniPHR against the original patient data. Also, with the patient data filled in the *open*EHR ontology, we employed SPARQL queries on the KB portion of the database to verify the populated data from OmniPHR compared with the original data.

Table 21 – Patient's health data from different sources.

Demographic			
Patient Number	Gender	Data of birth [1]	Language
NNN	-	YYYY-MM-DD	English
Admissions			
Source	Admission [1]	Discharge [1]	Type
*open*EHR	2157-01-11 16:56:00	2157-01-19 14:58:00	EMERGENCY
HL7/FHIR	2157-03-07 11:08:00	2157-03-10 13:50:00	ELECTIVE
Legacy	2157-11-17 12:11:00	2157-11-20 13:05:00	EMERGENCY
Blood pressure			
Source	Date [1]	Diastolic	Systolic
*open*EHR	2161-07-02 21:00:00	62	125
HL7/FHIR	2161-07-02 23:00:00	81	140
Legacy	2161-07-03 01:00:00	70	135
Diagnosis			
Source	Type [2]	Code	Description
*open*EHR	APR	NNNN	Spinal Procedures
HL7/FHIR	APR	NNNN	Other Compl. Treatment
Legacy	HCFA	NNNN	Septicemia
Heart rate			
Source	Date [1]	Frequency	
*open*EHR	2167-01-13 10:00:00	68	
HL7/FHIR	2167-01-13 11:00:00	71	
Legacy	2167-01-13 12:00:00	73	
Microbiology			
Source	Chart Time [1]	Item ID	Description
*open*EHR	2161-11-24 06:00:00	NNNNN	Blood Culture
HL7/FHIR	2161-11-26 20:15:00	NNNNN	Sputum
Legacy	2161-12-14 15:50:00	NNNNN	Urine
Prescriptions			
Source	Start Date [1]	End Date [1]	Drug Name Generic
*open*EHR	2157-01-11 00:00:00	2157-01-12 00:00:00	Fluconazole
HL7/FHIR	2157-01-12 00:00:00	2157-01-13 00:00:00	Heparin Sodium
Legacy	2161-07-07 00:00:00	2161-07-08 00:00:00	Potassium Chloride

[1] All data is anonymized.

[2] Types: HCFA (Health Care Financing Administration) and APR (All Payers Registry).

Source: Prepared by the author.

Following the *open*EHR and HL7/FHIR standards, we used anonymous patient data available in the MIMIC III database (JOHNSON et al., 2016) to submit to the model prototype. The version employed of MIMIC III is v1.4, with 58,000 subjects. Each patient was represented by the standards supported by the model, i.e. (a) *open*EHR, (b) HL7/FHIR and (c) legacy (MIMIC III). The MIMIC III database has real patient health data, although anonymized. In order for the tests to be the closest to reality, we use this data to compose the three formats supported by the model. In the case of the *open*EHR standard, we used the EHRServer (EHRSERVER, 2017) platform through the EHRCommiter component to populate the base of its database in PostgreSQL with data extracted from MIMIC III. This component was also used to extract the data in XML format. In that sense, we obtained a document in the *open*EHR format generated from a solution that admittedly follows this pattern. In the case of the HL7/FHIR standard, we performed the same script, using both the API and documentation in the XML format available on the FHIR website (FHIR, 2017). Finally, in the case of patient data that represented the legacy standard, data were extracted directly from the MIMIC-III base for plain text, in CSV (comma-separated values) format.

For instance, Table 21 shows a PHR used in the tests, with health data such as demographic, admissions, blood pressure, diagnosis, heart rate, microbiology and prescriptions of a patient, extracted from the different sources, but all of them originating from MIMIC-III database. The column 'Source' displays the source format used in the model. Some data are anonymized to do not identify the patient. The model evaluation was carried out using the Virtuoso database, populated through the OmniPHR with the archetype and ontology data. OmniPHR populates the relational portion of the Virtuoso database with the *open*EHR archetype data and the knowledge base portion with the *open*EHR ontology data. We employed SQL queries to check the consistency of the replicated data in OmniPHR against the original patient data. As well, with the patients' data filled in the *open*EHR ontology, we employed SPARQL queries on the KB portion of Virtuoso to verify the populated data from OmniPHR compared to the original data.

In the database, the inserts execute with the following SPARQL syntax:

```
insert in graph <openehr>
{ <http://subject> <http://predicate> <http://object> };
```

Moreover, we can use SPARQL queries in the normal way (W3C standard) or through SPASQL (SPARQL within SQL), which follow a pattern similar to SQL, with the syntax:

```
SPARQL SELECT * FROM <openehr> WHERE { ?Subject ?Predicate ?Object };
```

In this way, in the same database, we can verify the data with the same query pattern, in either SPARQL or SPASQL. Moreover, with data in both formats stored in the same database, we avoid possible integration problems between different databases. This feature proved to be efficient for verifying the inserted and updated data in relational and semantic formats. To collect the results, we ran SQL and SPARQL queries in Virtuoso to compare the conversions performed by the model, compared with the original data of the three formats. In the first execution of the selected data set, no manual interference or training adjustment was performed.

Only the relevant fields and values were determined. The accuracy achieved in the first run was on the order of 66%.

The accuracy represents the part of fields converted correctly, as expected, plus the fields that were not selected and not converted as expected, in relation to the total population of the records. The precision reached was 78.57%, i.e., the percentage of relevant fields that have been successfully converted. Recall achieved 74.32%, i.e., the percentage of fields that have been selected for conversion. Finally, the F1-score reached 76.39%, i.e., the harmonic mean of the precision and recall results. These numbers were initially lower than expected. To improve the failures of conversions on the first run, we had to perform additional training, improving the unfilled fields, as well as investigating the reasons.

Taking as sampling, consider the gender field. The name of this field is `gender` in HL7 FHIR and the values are defined by extensive writing (`male/female/other/unknown`). In MIMIC-III, this field has the same name, but the possible values are characters (`M/F`). However, in the *open*EHR standard, this field has several possibilities of filling because there is a dedicated archetype for this purpose. This archetype has several items, such as: administrative gender, legal gender, anatomical sex, gender expression, gender identity, and preferred pronoun.

7 DISCUSSION

This chapter presents the main findings found in the research, what are the limitations that the proposal has at the moment and, finally, the challenges and opportunities for future work.

7.1 Two-layer Architecture Proposal Analysis

Analyzing the Related Work on Chapter 4 and as we can see in Table 12, some models concentrate all patient's health data on single or multiple servers, following a centralized client-server architecture. Others models propose distributed architectures, although none currently uses P2P, just as future work in the case of HDEHR (XIA; SONG, 2012). Additionally, we analyzed models whether they provide security or privacy support for patients and health care providers use, where only few models tackle this subject. Finally, we researched what standards for interoperability the models support and only HDEHR (XIA; SONG, 2012) and CF (SAFAVI; SHUKUR, 2014) not specifically mention this subject.

Analyzing the results (summarized in Table 17), the 'Messages Present' column demonstrate an increasing number of transmission capacity and communication overhead. For instance, it is possible to compare the setup #3 with 400 nodes and setup #4 with 800 nodes, where the average number of messages present in the network and transmitted at the same time is doubled (6518 versus 13269 in the 'Test A', and 6593 versus 13457 in the Test #2), but the latency is very similar (0.479s versus 0.491s in 'Test A', and 0.403 versus 0.475), as well as the average number of hops (5.32 versus 5.54 in the 'Test A', and 5.34 versus 5.57 in the 'Test B'). Analyzing the other setups and tests, it is possible to observe that this behavior is maintained. Even with the number of nodes, routing overlays and backbone routers increasing, the number of messages being transmitted also increases, but the latency remains stable or even decreased in some cases. It is possible to observe, for example, that the latency in tests 1 and 2 with 1200 nodes (0.560 and 0.539, respectively) is much similar as in tests 1 and 2 with 3200 nodes (0.560 and 0.545, respectively). This demonstrates that the network topology, employing the Chord algorithm, was able to answer an increasing number of users and requests, but without increasing the delivery time significantly.

Another analysis is related to the results obtained from the parameters of tests 1 and 2. The Chord algorithm manages the nodes to be kept close. In this sense, it is possible to observe that in most results, the number of 'One-way hop count' increases slightly between tests 1 and 2 for the same setup. This happens due to the increase of routing overlays and backbones routers between the parameters of tests 1 and 2. However, latency remains stable or even decreased. This demonstrates that while increasing the number of routing overlays and backbone routers, for the same number of nodes, latency is not impacted. The tests also showed that there was no impact in performance with inputs and outputs of nodes, demonstrating the adequate capacity of elasticity and scalability of the P2P network following the Chord algorithm.

Regarding the standard for the health data proposed to be used in the OmniPHR model, to promote interoperability between different standards and among health care providers, the proposal is the use of the open standard *open*EHR. This standard is integrated with other standards specialized in specific health data types, such as laboratory exams results. Moreover, in order to enable proper distribution model, this open standard follows the principle of partitioning the PHR in datablocks. OmniPHR model proposes to distribute PHR in a P2P network, and this involves several challenges, such as rules to determine how to divide and replicate the datablocks. For this, the open standard that OmniPHR use divides PHR in structures of datablocks organized hierarchically. The evaluation sought to reflect the division, replication, and communication of datablocks in the network.

7.2 Performance Experiments Analysis

In the test scenarios, the number of users accessing the network was the number of concurrent sessions connected to the network, with the same increasing number of requests to the network (MORABITO, 2016). We chose a private Blockchain to restrict the management and access of network participants, thereby avoiding unauthorized sharing. This approach used mining resources and data evaluation more effectively by limiting access only to members of the network. In particular, evaluation in our private network was only performed by superpeers rather than burdening client nodes (which only produce and consume datablocks registered in the Blockchain). Two other factors justified our use of a private Blockchain network: (a) to facilitate the traceability of updates and (b) to reduce intermediaries in data exchanges since the superpeers concentrate the execution of operations on health records. Moreover, we applied the *open*EHR standard since it stored data in meta-data blocks, which integrates seamlessly into the Blockchain model. Our OmniPHR prototype accepts JSON and XML, though we applied XML predominantly within the Blockchain and for the evaluation tests since XSD is useful to evaluate content and typing.

This study just focused on private Blockchains instead of public Blockchains due to data security and privacy issues, as well as due to the specific domain of health care targeted by OmniPHR. Therefore, we did not allow access to other nodes since we handled sensitive health data that should only be shared by health providers and patients. Although there were some periods with communication problems in the network (i.e., some nodes were not accessible), these periods were generally short. Our Blockchain solution ensured that superpeers knew about the distribution of other nodes connected to them. In particular, since the Chord algorithm provided access to nodes with replicated content, superpeers could access other nodes with replicated data even though some nodes had communication problems. As a result, the overall operation of our solution was not impeded. Another aspect is regarding smart contracts used to evaluate the permissions granted on the PHR. The smart contract can specify who can access PHRs and what permissions each client can get on the data. Therefore, a smart contract on the

OmniPHR prototype maintains the security and privacy of health records.

One difficulty faced in evaluating the OmniPHR prototype stemmed from the challenge of submitting data to the model. To test the prototype, we had to submit a considerable volume of health records to evaluate its performance. However, the results from the load tests shown in Figure 23 indicated that in the heavy scenario response times stabilize around 500ms. In general, the OmniPHR prototype demonstrated average responses below one second. Although average response times grew with the load and number of users, response times remained low even as the loads increased. In particular, response times are nearly instantaneous with smaller loads and few simultaneous accesses. The network still responded quickly, however, even with larger simultaneous loads and accesses.

Comparing our performance experiments, Table 13 summarizes results obtained by related work. Although these studies espouse the benefits of applying Blockchain technologies to the health care domain through qualitative evaluations, few studies present empirical results to substantiate their claims. Hence, we focus on qualitative analyses that evaluate the performance and efficacy of integrating health records via Blockchain technologies. Although all projects use some Blockchain technology in their implementations, only Healthbitt (RONO, 2016), MedRec (EKBLAW, 2017), Patientory (MCFARLANE et al., 2017) and FHIRChain (ZHANG et al., 2018) applied at least one health data standard and focus on providing access to both health providers and patients. Among the related work efforts presented in Table 13, seven used at least one of the two cross-industry platforms: Ethereum or Hyperledger. Most of these studies used Ethereum (LAZAROVICH, 2015; OUADDAH; ABOU ELKALAM; AIT OUAHMAN, 2016; EKBLAW, 2017; MCFARLANE et al., 2017; DAGHER et al., 2018; ZHANG et al., 2018) as their Blockchain platform and only one used Hyperledger Fabric (ICHIKAWA; KASHIYAMA; UENO, 2017). The Ethereum platform uses the Ether (ETH) crypto-currency, whereas Hyperledger is not associated with any crypto-currency.

Related work focuses largely on describing how models can utilize Blockchain technologies (YUE et al., 2016; LAURA; KOIVUMÄKI; SARANIEMI, 2016; ICHIKAWA; KASHIYAMA; UENO, 2017; HENZE et al., 2017; XIA et al., 2017). In contrast, our research presented in this study focuses on demonstrating the viability of Blockchain technologies by evaluating the behavior of the OmniPHR prototype in production health record scenarios. Moreover, unlike related work that use conventional Blockchain platforms like Ethereum or Hyperledger, OmniPHR uses the Chord algorithm, which supports replication. Conventional Blockchain platforms generally follow the original Blockchain concept applied to crypto-currencies, which replicate data to all nodes in the network. In contrast, the Chord replication algorithm enables finer-grained control over how much, how, and where to replicate the data, thereby enabling more granular control of replications. Our results show that Chord optimizes performance, although data redundancy is reduced. In addition, by storing datablocks in ontology format, i.e., in the Ontology Web Language (OWL), the KnowledgeBase enables the creation of semantic rules that allow inferences about possible patient health problems.

7.3 Multi-Blockchain Proposal Analysis

Analyzing the results obtained with Multi-Blockchain proposal, we can check that the first executions take extra time. This time we observe due to the initialization of configurations that the Blockchains network needs to perform, such as the knowledge of the middleware and the other nodes location in the network. For the other executions, we can observe that the performance improves significantly. We can also note that insert operations take a little longer than query-only operations. Another observation is that the operations within the internal Blockchain are faster, compared to the hop needed to reach other Blockchains. However, we noted that the query times of one block are less than one second. In addition, we observed that performing batch operations like 100, 500, and 1000 transactions are close, with a slight deterioration of performance (THAKKAR; NATHAN; VISWANATHAN, 2018). The results showed that the insert operations of blocks from the health provider performed in the worst cases under 9 seconds. In the case of queries of data records in the internal Blockchain, the performance was almost in real-time. The situations in which the performances have degraded somewhat were due to the search of data in another Blockchain with a hop, but even so in low times.

Comparing our architectural proposal and the results obtained with the related work seen in Table 14, we can visualize three differences, as follows.

(i) The first observation is that, among these studies, only Hawig et al. (HAWIG et al., 2019) proposes the use of a health data standard, in the case HL7 FHIR (SARIPALLE; RUNYAN; RUSSELL, 2019). The other works do not address and do not specify interoperability standards for health data. Our work proposes, as a first option, the use of the *open*EHR (YANG; HUANG; LI, 2019) health data standard, due to its integration with other standards, such as HL7 FHIR (SARIPALLE; RUNYAN; RUSSELL, 2019), DICOM (GONÇALVES-FERREIRA et al., 2019), SNOMED-CT (TEODORO et al., 2018) and LOINC (WULFF et al., 2018), in addition to the standard blocks through the archetypes format. Nevertheless, our model can also support the HL7 FHIR format since the proposal includes the transformation of this format into *open*EHR (ROEHRS et al., 2018).

(ii) The second difference concerns the implementation framework, where some studies propose the use of Ethereum (ERHARDT et al., 2017; SHARMA; CHEN; PARK, 2017; GOERTZEL et al., 2017; VO et al., 2018; WANG; ZHANG; KIM, 2018), others Hyperledger (ERHARDT et al., 2017; VO et al., 2018; SHE et al., 2019) and others do not mention a specific framework. In this case, our proposal is based on the Hyperledger framework, to facilitate the implementation of our Multi-Blockchain model, due to some important characteristics such as:

(a) being an open-source framework, allowing debugging and behavior analysis,

(b) the same way works with open and popular programming languages,

(c) that allows the creation of private Blockchains, important in the case of sensitive health data and (d) without the connection with some crypto.

(d) without the connection with some crypto (THAKKAR; NATHAN; VISWANATHAN, 2018).

(iii) The third observation is regarding to the type of health data stored in the Blockchain, where some studies mention the use of EHR (SHARMA et al., 2018; MA; SHI; LI, 2019; HAWIG et al., 2019), others mention PHR (ERHARDT et al., 2017; SHARMA; CHEN; PARK, 2017; SHE et al., 2019) and the others do not specifically mention some type of health data. In this case, our proposal aims to attend the patient emphasizing mobility, i.e., the possibility of the patient having, in fact, under control their personal health data, composing the concept of PHR.

Finally, comparing the obtained results, as a main difference, we seek to demonstrate the ability of our model to scale the patient's health data in a distributed, interoperable, and standardized architecture.

7.4 Semantic Interoperability Proposal Analysis

To maintain compatibility between *open*EHR and the HL7 FHIR standard, we filled the fields that are compatible between the two standards. In this case, the compatible type is the administrative gender field. According to the *open*EHR documentation: "This aligns with HL7 FHIR 'Person.gender'" (INFORMATICS, 2018). We trained the OmniPHR model based on the possible *open*EHR definitions and values. In this case, the first time the prototype ran, it did not populate the gender field completely because the possible types and possible values in the source were different. After training the network, stating that `gender` could receive characters as abbreviations, we re-executed the model, and OmniPHR recognized these types populating the ontology correctly. Another example, in the case of unstructured data, is related to allergies. In MIMIC-III allergies are described in a descriptive text field, e.g., Allergies: "Codeine and shellfish", or Allergies: "Codeine/Ambien/Shellfish Derived", or Allergies: "He has an allergy to CODEINE,...". However, in *open*EHR, it is a structured field forming an adverse reaction list. With the help of NLP, OmniPHR can extract the relevant data from the sentences and convert this unstructured information to the *open*EHR structured list.

We retrained the model with the new possible values for the unconverted parts. Specifically, in the case of the GATE platform, it provides an API with a series of machine-learning components (CUNNINGHAM et al., 2013), which we used to recognize the patterns sent according to the standards adopted in *open*EHR. Then, OmniPHR could convert more data sent to the tests, returning a larger number of expected results in the *open*EHR format. After the tuning and training, the accuracy scores improved and reached 81%, precision 87.34%, recall 88.46%, and the F1-score 87.9%.

Revisiting the related work and analyzing the results obtained for comparison, it was difficult to draw a parallel. Likewise, it was difficult to find studies that used a metric of statistical analysis to verify, quantitatively, or qualitatively, the treatment of the health records' interoperability. Thus, to obtain a more accurate evaluation of the model, we used the F1-score metric. In this way, we obtained a parameter for the algorithm regarding the precision, recall, and accuracy. Initially, we observed that the algorithm obtained low precision, recall, and accuracy. Lower than we expected. However, with machine-learning training, we observed that there was an improvement.

In addition to the low score in the first run, a manual intervention effort was necessary to correct and train the model to improve the results. However, there is the advantage of the ability to reuse training learned in new executions, extracting data from structured or unstructured fields. Finally, the model was limited to tests with the MIMIC-III database and using the English language. Thus, as future directions, the model needs to be evaluated using other databases, following other health standards and tested using other languages.

7.5 Research Limitations

This research is limited to aspects related only to PHRs rather than also including EHRs or electronic medical records, for example. In this sense, the review focused exclusively on articles addressing the inherent PHR concepts. This research sought to answer the research questions that were proposed in order to obtain an outline of the current literature related to PHRs without specifically assessing any computer system that refers to the use of PHR. The research was limited to obtaining articles published in a number of scientific portals related to ICT and health. Our research was reduced to studies found from these websites when we implemented the steps of the SLR methodology. We focused our work on scientific articles and did not address commercial or more technological approach solutions.

A limitation of the model concerns the type and location of the data. First, the data must follow the standards supported by OmniPHR model and be located in the model-enabled paths. In this way, the architecture of the model is able to access and maintain the data. This means that a patient's data that is not in the scope of the model will not be part of the sharing, either with the patient or with other health organizations. Besides, as the premise is that each datablock must have a responsible, the model needs to be able to determine the author of each data, whether patient, health care provider or sensor, ensuring the authorship of each datablock. For instance, the demographic datablock is the patient's responsibility, while health care providers are responsible for other datablocks such as diagnosis.

In addition, the model needs to store data on the node closest to the user, with copies in other nodes. For instance, datablocks created by a physician in a hospital are stored in the datacenter of health organization. Similarly, data reported by patients are stored on the routing overlay with copies in other nodes. In this way, another limitation is that, as default, data are shared only between health care provider and patient. This means that for a health organization to access the patient's data in another one, the patient must authorize this access.

Another potential problem that the model should deal with is the possibility of occurring duplicate data entry, as it usually happens in health organizations (MCCOY et al., 2013). This problem may occur mainly with patients registration data, e.g., when a patient is admitted in a health organization, and the registration is not found. Another possibility is with legacy data, when a health organization, that already has the PHR, wants to join the system. To avoid this problem, the system needs to identify the patient unequivocally, leaving no doubt. OmniPHR provides a mechanism to generate a single hash code to identify patients, following the *open*EHR standard for this identifier (OPENEHR, 2017).

PHR can be composed of many datablocks throughout the patient's lifetime and also by a large number of attachments. Examples of attachments would be the laboratory exams results and medical images. These images usually have relatively large sizes, up to several tens of megabytes (HUSSEIN; BADR, 2013). As health records can be formed by many data, the intent is that queries are not made of all health records at once but in parts. In addition, these health records can be divided in order to be sought only the most recent data in a paginated format. An example would be the case of laboratory exams results returned in date order. That is, from the most recent to the oldest, using pagination. So, just return the latest laboratory exams results and, if necessary, then seek the older data in another query. This mechanism provides optimization of database queries because records are always divided into pages, generating less traffic on the network.

One of the major difficulties in promoting the interoperability of health records is dealing with a range of different data types and content variability of EHRs, not only considering their structure but also considering that many data are stored in textual format inside descriptive fields. The experiments on OmniPHR were limited to the data and types provided by MIMIC-III, as well as patient data and data types provided by *open*EHR and HL7 FHIR. Considering the effort involved in training, we believe that it would be possible to extend it to other formats because *open*EHR can be translated into other languages that follow the ISO 639-1 standard; however, further experimentation is needed.

Our performance experiments did not cover the execution of business rules and inferences about records, such as specific evaluations of the content of patients' health records. Instead, we limited our OmniPHR prototype to joining datablocks that formed a unified view of patient data. In particular, our load tests only focused on evaluating the distribution and traffic of the blocks of records based on Blockchain technology and the *open*EHR standard. We made this provision for isolating the performance evaluation of the Blockchain solution without the interference of the usual business rule validations that HIS have.

Another limitation is for image files, such as DICOM images. These images can occupy large spaces because of their size in megabytes. Replication of these files in the Blockchain is not foreseen, although the location address is provided. In this way, the images are stored off-chain with a content hash code, and only the address where the images are located is replicated to the network. We created the test scenarios in order to stress the system and verify that it remained stable without generating errors or crashes, such as OOM (Out Of Memory). We went to the limit of having at least one block of data from each registered patient handled concurrently. We tried to verify if the system remained stable of the original form as it was constructed, without using a special tuning of optimization.

Analyzing the results, it is important to highlight as limitations of the aspects of the tests related to possible technological alternatives, which we can use to improve the performance, as parallelism in the processing, use of cache database, and improve on the initializing operation (THAKKAR; NATHAN; VISWANATHAN, 2018). Another limitation was that we tested the performance while the model was responding, i.e., without producing timeout or memory overflow errors.

7.6 Challenges and Opportunities

An important challenge for the model is to guarantee the identity and authenticity of the informant, whether patient, physician or a sensor connected to the patient. OmniPHR predicted a security module in its architecture and is based on the structure of proposed distribution and security mechanism of datablocks distributed in a P2P network, with encryption and the digital signature of datablocks to ensure the authenticity. This aims at ensuring the health records chain validity and data privacy.

Nevertheless, attackers can try to spoof these parts, try to decrypt parts, try to gain access to other nodes, and try to reassemble datablocks (TAPSCOTT; TAPSCOTT, 2016). While Blockchain technology helps prevent datablock fraud, it remains a challenge to ensure that only authentic informants can access the health records. Although the model demonstrates the potential to preserve the privacy of patients, further testing for security and privacy is required.

In relation to the architecture model, a component that needs future work is the Translator Component of the Datablock and Service Module. This component aims to convert and equalize communications with heterogeneous health systems. The component proposes to convert data coming from open or proprietary standards, and for this purpose, OmniPHR should deal with the possibility of integration using open standards such as HL7/FHIR and *open*EHR or an equivalent ontology in case of integration with proprietary systems, which is another challenge.

Many health providers adopt their own formats for the use of health records, and even when using open standard usually do not share them with other organizations. Thus, a patient may have health records scattered in several organizations. With OmniPHR, health care providers can be able to have access to complete PHR of patients assisted since the first contact with the health organization. Health organizations already have a cost of maintaining medical records (LI et al., 2013) and not integrated with other institutions. The benefit of our model to providers is to have their patient's health data, that is already stored in the organization, always up-to-date, beyond the possibility of extracting medical statistics to improve the quality of care.

However, in this sense, there are several questions and challenges to face, such as:

(a) Would it be necessary for all data to be shared online, or could it be according to a configurable periodicity and at idle moments?

(b) Is patient's data reported by a health care provider and shared with another one available indefinitely or for a fixed time?

(c) When sharing data, are all patient's data available by default (clinical and administrative) or should the patient select which ones to share?

(d) In case the patient needs to select, how can patient accomplish this task without great knowledge?

8 CONCLUSIONS

This study started with the purpose of to raise and discuss the main issues regarding PHRs and identify the concepts of technology in this area. To answer the research questions in this paper, we sought first to systematize and qualify the information that served as a source for the survey. Aside from answering all the specific research questions and relating them in the taxonomy, we can also rank the PHR with regard to goals, negotiation types, and architectures. The answers and classifications obtained contribute to the achievement of a coverage degree of searches that are identified in various aspects regarding the PHR. The physician-patient relationship traditionally consists of total dependence of the patient on the physician. In addition, the fragmented nature of the health system can impose a costly burden on physicians. The PHR can be a solution to this problem, although obstacles still persist, including support for reaching this paradigm, where the ownership of the data belongs to the patient.

In this work, we presented a distributed architecture proposal named OmniPHR. This solution seeks to address recurrent needs in the adoption of PHR by patients and health care providers. The OmniPHR purpose consists of partitioning PHR in datablocks distributed on a P2P network. Thereby OmniPHR maintains characteristics of datablocks distribution having spread copies of these parts on the network. The user can access PHR data through different devices. Consequently, OmniPHR is a mobile-health model which uses the diversity of computing devices connected to the patients or to the environment where they are inserted at any time, to be part of a collaborative and distributed network. The PHR data appear to be centralized from the logical viewpoint of patient and health care provider, but in fact, PHR is physically decentralized. This model proposes an architecture for users to obtain a single view of patient health records with scalability, elasticity, and interoperability. With the PHR data scattered in several health organizations where patients had contact, OmniPHR proposes to mitigate many problems and barriers in the adoption of PHR providing a unified viewpoint of PHR. The model aims to support patients to take advantage of having their health history single, as well for health care providers have their patients' health data up-to-date. Hence, OmniPHR proposes a model where everyone involved has benefits sharing health records.

There are several health standards for PHR use. Revisiting the proposal of scientific contribution, this research aimed to present a model to promote the interoperability between different standards with semantic capability. We identified many ways to promote interoperability between health records, such as the use of archetypes, metadata, ontologies, software agents, templates, and terminologies. To ensure interoperability, the model selected the *open*EHR reference model as the component centralizer. We proposed in OmniPHR the use of NLP as a CNL component to help automate the conversion process. The results obtained with the evaluation of the prototype were promising, demonstrating the feasibility of the OmniPHR model using the health records from the MIMIC-III database. OmniPHR demonstrated compliance with the requirements of semantic interoperability and unified view of patient data.

The results of this contribution demonstrated the possibility of obtaining a unified and up-to-date view of health data, presenting a solution based on artificial intelligence with NLP, ontology, and an open health standard to achieve semantic interoperability. In addition, OmniPHR presented benefits, such as the possibility of obtaining inferences about the patients' health. In future work, we intend to evaluate the model with data from larger health databases and focus on increasing the possibilities of benefits for patients and health care providers. Other important aspects to discuss are data distribution, scalability, security, and privacy. In addition, the prototype can expand to integrate with other open and proprietary health standards.

This work presented the prototype implementation and evaluation of the OmniPHR architecture model that integrates distributed health records using Blockchain technology and the *open*EHR interoperability standard. The OmniPHR prototype comprises a novel Blockchain-based design that optimizes health data replication across computing nodes. We evaluated the performance of our OmniPHR prototype by subjecting it to loads of thousands of concurrent sessions transmitting datablocks on a network of 10 superpeers. We also evaluated implementation strategies related to the replication of health-oriented Blockchain solutions to promote the unification of patient health data.

We started this study with the objective of contributing with a differentiated proposal of architecture to support the distribution of patient health data. We can observe the growth of Blockchain technology in supporting the use of patient health records. However, the traditional Blockchain proposition tends to replicate the data blocks for all nodes, which in terms of health data and for a large number of integrated health organizations, can be a problem due to the size of each medical records may have. After presenting the essential background for the fulfillment of our proposal, we presented the OmniPHR Multi-Blockchain model in a different format from the one we presented in previous works (ROEHRS; COSTA; ROSA RIGHI, 2017; ROEHRS et al., 2019). In the following, we presented the methodology and related work. In this sense, we researched related work close to our proposal, which aimed to address the challenges with the health records distribution through innovative architectures.

Finally, we presented the obtained results, and we compared them with the existing proposals, where we can notice the differences in the proposals. We highlighted our proposal of distributed Multi-Blockchain architecture with the help of the Hyperledger framework, the use of *open*EHR health data standard, and focus on PHR. The results demonstrated the potential of the OmniPHR Multi-Blockchain model and the need for greater testing in different usage scenarios.

8.1 Contributions

For the completion of the work, we were able to identify and propose a broad taxonomy for the scope of work, which was created after an analysis of the relevant articles in the last decade. In the taxonomy, we were able to identify and group a number of types and PHR classifications ranging from "Structures" and types associated with "Functions" to the types of "Architectures" applied to PHRs. Having established the taxonomy, we observed other important relationships to understand PHRs. We noticed aspects regarding concerns and challenges in the adoption of PHRs as well as the main data types.

In addition, we were able to identify several standards regarding PHR, where it was possible to verify those that were most important in the current scenario. Regarding user profiles, we identified the main users representing these types of profiles, as well as their responsibilities when they access PHRs. We were able to identify the techniques and methods used in the input of information into PHRs.

The scientific contribution of this work is to present a different proposal of distribution and interoperable architecture for PHR. The evaluation of the model demonstrates that OmniPHR is able to promote PHR divided into datablocks and proportional distribution in a routing overlay network through the Blockchain technology. The results showed that even increasing the number of nodes, and consequently obtaining a larger number of messages being transmitted at the same time in the network, the latency remains stable. This demonstrates that OmniPHR is able to support a growing number of nodes and requests without increasing the delivery time significantly.

In order to better promote the interoperability between the different existing standards and to add semantic capacity, this study aimed to present an interoperable and semantic model. To meet interoperability, the model selected the *open*EHR reference model as the component centralizer. As a conversion helper, the model proposed the use of NLP as a controlled natural language conversion component. And, as a repository, the model proposed the use of the multi-model database, supporting both relational data and RDF triple store.

The following are a summary of the lessons learned from conducting our research:

(a) Combining the *open*EHR standard with Blockchain technologies created a unified and interoperable view of health data. Even with some limitations, such as not executing business rules on the prototype (since it is not a complete system), we observed promising results of the architectural model using our private Blockchain platform.

(b) Applying the Chord algorithm for directed and limited data replication is a more scalable alternative than conventional crypto-currency platform replication models, where all nodes receive all data. Chord's scalability is a critical factor to support health data effectively. In particular, it enables data replication with restricted access, providing control and management by patients and health care professionals.

(c) The results of our empirical evaluations showed that the OmniPHR Blockchain architecture provided adequate network-level performance. It, therefore, appears that patient health records can be integrated effectively via a Blockchain network using technologies applied to the treatment of large masses of data and an interoperable health data standard.

As additional contributions, four articles were published and made available to the academic community, according to the following list. One article was published in the Journal of Medical Internet Research (JMIR), one article was published in the IEEE Journal of Biomedical and Health Informatics (JBHI), and two in the Journal of Biomedical Informatics (JBI):

1. Roehrs A, da Costa CA, da Rosa Righi R, de Oliveira KSF. **Personal Health Records: A Systematic Literature Review.** Journal of Medical Internet Research (JMIR) 2017; 19(1):e13. DOI: 10.2196/jmir.5876. PMID: 28062391. PMCID: 5251169. URL: https://www.jmir.org/2017/1/e13.

2. Roehrs A, da Costa CA, da Rosa Righi R. **OmniPHR: A distributed architecture model to integrate personal health records.** Journal of Biomedical Informatics. Volume 71. 2017. Pages 70-81. ISSN 1532-0464. DOI: 10.1016/j.jbi.2017.05.012. URL: http://www.sciencedirect.com/science/article/pii/S1532046417301089.

3. Roehrs A, da Costa CA, da Rosa Righi R, Rigo SJ, Wichman MH. **Toward a Model for Personal Health Record Interoperability.** IEEE Journal of Biomedical and Health Informatics. 2018 May. 14; Volume 23. Issue 2. Pages 867-73. DOI: 10.1109/JBHI.2018.2836138 URL: https://ieeexplore.ieee.org/abstract/document/8358689.

4. Roehrs A, da Costa CA, da Rosa Righi R, da Silva VF, Goldim JR, Schmidt DC. **Analyzing the performance of a blockchain-based personal health record implementation.** Journal of biomedical informatics. 2019 Mar 4:103140. DOI: 10.1016/j.jbi.2019.103140 URL: https://www.sciencedirect.com/science/article/pii/S1532046419300589

In addition to these publications, during the development of this study, there were some indirect publications, as co-author:

1. Wichman MH, da Costa CA, Roehrs A, Bandeira D, da Rosa Righi R. **Integration Between Electronic Health Records Standards Using Ontologies And Rules.** IADIS International Conference Applied Computing. 2017. Pages 23-30.

2. Quaini T, Roehrs A, da Costa CA, da Rosa Righi R. **A Model For Blockchain-Based Distributed Electronic Health Records.** IADIS International Journal on WWW/Internet. 2018 Jul 1;16(2).

3. Montenegro JLZ, da Costa CA, da Rosa Righi R, Roehrs A. **A Proposal for Postpartum support based on Natural Language Generation Model.** In: 5th Annual Conf. on Computational Science & Computational Intelligence (CSCI'18), 2018, Las Vegas. Proceedings of 5th Annual Conf. on Computational Science & Computational Intelligence (CSCI'18). New York: IEEE, 2018. v. 1. p. 1-756.

8.2 Future work

In future studies, we envision a focus on the challenges and issues related to security, privacy, and trust, which directly affect the users' confidence in adopting the PHR. Although these questions have existed for a long time, they do not have definitive answers yet. Other aspects that can be studied and that are important to improving the user experience are questions about usability, personalization, familiarity, and comfort. Another aspect that can serve as a future study is to explore the models of architecture and the implementation of PHR following the expansion of the use of technologies such as wearable computing, IoT, and artificial intelligence that are applied to health.

As future work, the model needs more evaluations, mainly regarding security, privacy, and integration with other systems. Moreover, as can be seen, there are several challenges to be worked on and answered, ranging from decisions to flexible the model regarding access rules and data replication, to subjective questions that arise, likewise how patients can manage and share their data in practice.

In addition, as future study, we intend to focus on increasing the possibilities of benefits for patients and health care providers. In addition, the prototype can be expanded to integrate other health data standards. Other important aspects to be discussed are related to the distribution of data, as well as data security and privacy.

In future work, we plan to evolve our OmniPHR prototype to incorporate additional databases and conduct additional tests to evaluate its performance in even more scalable and realistic production environments. Other evaluations we plan to conduct involves data security and privacy, especially in the case of external access to private Blockchain networks.

REFERENCES

ADUFU, T.; CHOI, J.; KIM, Y. Is container-based technology a winner for high performance scientific applications? In: NETWORK OPERATIONS AND MANAGEMENT SYMPOSIUM (APNOMS), 2015 17TH ASIA-PACIFIC, 2015. **Anais...** [S.l.: s.n.], 2015. p. 507–510.

AGARWAL, R. et al. If we offer it, will they accept? factors affecting patient use intentions of personal health records and secure messaging. **Journal of medical Internet research**, [S.l.], v. 15, n. 2, 2013.

AHMED, E. et al. Seamless application execution in mobile cloud computing: motivation, taxonomy, and open challenges. **Journal of Network and Computer Applications**, [S.l.], v. 52, p. 154–172, 2015.

ALABBASI, S. et al. Data types managed database design for dynamic content: a database design for personal health book system. In: TENCON 2014-2014 IEEE REGION 10 CONFERENCE, 2014. **Anais...** [S.l.: s.n.], 2014. p. 1–5.

ALIAKBARPOOR, Y.; COMAI, S.; POZZI, G. Designing a hl7 compatible personal health record for mobile devices. In: RESEARCH AND TECHNOLOGIES FOR SOCIETY AND INDUSTRY (RTSI), 2017 IEEE 3RD INTERNATIONAL FORUM ON, 2017. **Anais...** [S.l.: s.n.], 2017. p. 1–6.

ALTEROVITZ, G. et al. Smart on fhir genomics: facilitating standardized clinico-genomic apps. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 6, p. 1173–1178, 2015.

ALYAMI, M. A.; SONG, Y.-T. Removing barriers in using personal health record systems. In: COMPUTER AND INFORMATION SCIENCE (ICIS), 2016 IEEE/ACIS 15TH INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–8.

ANCKER, J. S.; SILVER, M.; KAUSHAL, R. Rapid growth in use of personal health records in new york, 2012–2013. **Journal of general internal medicine**, [S.l.], v. 29, n. 6, p. 850–854, 2014.

ANDREU-PEREZ, J. et al. Big data for health. **IEEE journal of biomedical and health informatics**, [S.l.], v. 19, n. 4, p. 1193–1208, 2015.

ARCHER, N. et al. Personal health records: a scoping review. **Journal of the American Medical Informatics Association**, [S.l.], v. 18, n. 4, p. 515–522, 2011.

ASC. **Accredited standards committee (asc) x12n insurance subcommitee**. Last accessed: 2019-04-02, Available from: `http://www.x12.org/x12org/subcommittees/asc-x12-rosters.cfm?strSC=N`.

ASPLUND, M.; LOVHALL, J.; NADJM-TEHRANI, S. In-store payments using bitcoin. In: NEW TECHNOLOGIES, MOBILITY AND SECURITY (NTMS), 2018 9TH IFIP INTERNATIONAL CONFERENCE ON, 2018. **Anais...** [S.l.: s.n.], 2018. p. 1–6.

ATTEMPTO. **Attempto project - attempto controlled english (ace)**. Last accessed: 2019-06-14, Available from: `http://attempto.ifi.uzh.ch/site/`.

AZARIA, A. et al. Medrec: using blockchain for medical data access and permission management. In: INTERNATIONAL CONFERENCE ON OPEN AND BIG DATA (OBD), 2016., 2016. **Anais...** [S.l.: s.n.], 2016. p. 25–30.

AZEVEDO, R. R. de et al. Towards a framework for ontology learning from interactions in natural language and reasoning. In: ANNUAL INTERNATIONAL CONFERENCE ON COMPUTER SCIENCE AND SOFTWARE ENGINEERING, 24., 2014. **Proceedings...** [S.l.: s.n.], 2014. p. 120–132.

BAIRD, A.; NORTH, F.; RAGHU, T. Personal health records (phr) and the future of the physician-patient relationship. In: CONFERENCE, 2011., 2011. **Proceedings...** [S.l.: s.n.], 2011. p. 281–288.

BANERJEE, A.; GUPTA, S. K. Analysis of smart mobile applications for healthcare under dynamic context changes. **IEEE Transactions on Mobile Computing**, [S.l.], v. 14, n. 5, p. 904–919, 2015.

BASKERVILLE, R.; PRIES-HEJE, J.; VENABLE, J. Soft design science methodology. In: OF THE 4TH INTERNATIONAL CONFERENCE ON DESIGN SCIENCE RESEARCH IN INFORMATION SYSTEMS AND TECHNOLOGY, 2009. **Proceedings...** [S.l.: s.n.], 2009. p. 9.

BELLAMY, J. Apta physical therapy outcomes registry data published in worldwide logical observation identifiers names and codes (loinc) database. **American Physical Therapy Association**, [S.l.], 2016.

BENDER, D.; SARTIPI, K. Hl7 fhir: an agile and restful approach to healthcare information exchange. In: COMPUTER-BASED MEDICAL SYSTEMS (CBMS), 2013 IEEE 26TH INTERNATIONAL SYMPOSIUM ON, 2013. **Anais...** [S.l.: s.n.], 2013. p. 326–331.

BENSON, T.; GRIEVE, G. **Principles of health interoperability**: snomed ct, hl7 and fhir. [S.l.]: Springer, 2016.

BHARTIYA, S.; MEHROTRA, D.; GIRDHAR, A. Issues in achieving complete interoperability while sharing electronic health records. **Procedia Computer Science**, [S.l.], v. 78, p. 192–198, 2016.

BIOLCHINI, J. et al. Systematic review in software engineering. **System Engineering and Computer Science Department COPPE/UFRJ, Technical Report ES**, [S.l.], v. 679, n. 05, p. 45, 2005.

BLAKE, M. B. An internet of things for healthcare. **IEEE Internet Computing**, [S.l.], v. 19, n. 4, p. 4–6, 2015.

BOSSEL, H. **Modeling and simulation**. [S.l.]: Springer-Verlag, 2013.

BOURGEOIS, F. C.; NIGRIN, D. J.; HARPER, M. B. Preserving patient privacy and confidentiality in the era of personal health records. **Pediatrics**, [S.l.], v. 135, n. 5, p. e1125–e1127, 2015.

BOURI, N.; RAVI, S. Going mobile: how mobile personal health records can improve health care during emergencies. **JMIR mHealth and uHealth**, [S.l.], v. 2, n. 1, 2014.

BRENNAN, P. F.; DOWNS, S.; CASPER, G. Project healthdesign: rethinking the power and potential of personal health records. **Journal of biomedical informatics**, [S.l.], v. 43, n. 5, p. S3–S5, 2010.

BRICON-SOUF, N.; NEWMAN, C. R. Context awareness in health care: a review. **international journal of medical informatics**, [S.l.], v. 76, n. 1, p. 2–12, 2006.

BUTLER, J. M. et al. Understanding adoption of a personal health record in rural health care clinics: revealing barriers and facilitators of adoption including attributions about potential patient portal users and self-reported characteristics of early adopting users. In: AMIA ANNUAL SYMPOSIUM PROCEEDINGS, 2013. **Anais...** [S.l.: s.n.], 2013. v. 2013, p. 152.

CACHIN, C.; VUKOLIĆ, M. Blockchain consensus protocols in the wild. **arXiv preprint arXiv:1707.01873**, [S.l.], 2017.

CAHILL, J. E.; GILBERT, M. R.; ARMSTRONG, T. S. Personal health records as portal to the electronic medical record. **Journal of Neuro-oncology**, [S.l.], v. 117, n. 1, p. 1–6, 2014.

CALIGTAN, C. A.; DYKES, P. C. Electronic health records and personal health records. In: SEMINARS IN ONCOLOGY NURSING, 2011. **Anais...** [S.l.: s.n.], 2011. v. 27, n. 3, p. 218–228.

CANTOR, M. N.; THORPE, L. Integrating data on social determinants of health into electronic health records. **Health Affairs**, [S.l.], v. 37, n. 4, p. 585–590, 2018.

CASTILLO, V. H.; MARTÍNEZ-GARCÍA, A. I.; PULIDO, J. A knowledge-based taxonomy of critical factors for adopting electronic health record systems by physicians: a systematic literature review. **BMC medical informatics and decision making**, [S.l.], v. 10, n. 1, p. 60, 2010.

CCR. **Standard specification for continuity of care record**. Last accessed: 2019-04-02, Available from: `https://www.astm.org/Standards/E2369.htm`.

CEN. **European committee for standardization - cen/tc 251 - health informatics**. Last accessed: 2019-04-02, Available from: `https://standards.cen.eu/`.

CHEN, R. Current challenges of ehrs for oncologists. **Oncology Times**, [S.l.], v. 38, n. 16, p. 1–10, 2016.

CHEN, Y. et al. Blockchain-based medical records secure storage and medical service framework. **Journal of medical systems**, [S.l.], v. 43, n. 1, p. 5, 2019.

CHIAUZZI, E.; RODARTE, C.; DASMAHAPATRA, P. Patient-centered activity monitoring in the self-management of chronic health conditions. **BMC medicine**, [S.l.], v. 13, n. 1, p. 77, 2015.

CHRISCHILLES, E. A. et al. Personal health records: a randomized trial of effects on elder medication safety. **Journal of the American Medical Informatics Association**, [S.l.], v. 21, n. 4, p. 679–686, 2014.

CHUNG, L. et al. **Non-functional requirements in software engineering**. [S.l.]: Springer Science & Business Media, 2012. v. 5.

COSTA, C. A. da et al. Internet of health things: toward intelligent vital signs monitoring in hospital wards. **Artificial Intelligence In Medicine**, [S.l.], v. 89, p. 61–69, 2018.

COULOURIS, G. et al. **Distributed systems**: concepts and design. [S.l.]: Boston: Addison-Wesley, 2012.

COULOURIS, G. F. et al. **Distributed systems**: concepts and design (5th edition). [S.l.]: Pearson Education, 2011.

CTAKES. **Apache ctakes – clinical text analysis knowledge extraction system**. Last accessed: 2019-06-14, Available from: `http://ctakes.apache.org/`.

CUNNINGHAM, H. et al. Getting more out of biomedical documents with gate's full lifecycle open source text analytics. **PLoS computational biology**, [S.l.], v. 9, n. 2, p. e1002854, 2013.

CUNNINGHAM, J.; AINSWORTH, J. Enabling patient control of personal electronic health records through distributed ledger technology. **Stud Health Technol Inform**, [S.l.], v. 245, p. 45–48, 2018.

CZAJA, S. J. et al. The usability of electronic personal health record systems for an underserved adult population. **Human factors**, [S.l.], v. 57, n. 3, p. 491–506, 2015.

DAGHER, G. G. et al. Ancile: privacy-preserving framework for access control and interoperability of electronic health records using blockchain technology. **Sustainable Cities and Society**, [S.l.], v. 39, p. 283–297, 2018.

DAVIS, B. P. **On applying controlled natural languages for ontology authoring and semantic annotation**. 2013. Tese (Doutorado em Ciência da Computação) — , 2013.

DHILLON, V. Designing decentralized ledger technology for electronic health records. **Telehealth and Medicine Today**, [S.l.], v. 2, n. 6, 2017.

DICOM. **Digital imaging and communications in medicine**. Last accessed: 2019-04-02, Available from: `http://dicom.nema.org/`.

DOLIN, R. et al. Health level seven interoperability strategy: big data, incrementally structured. **Methods of information in medicine**, [S.l.], v. 54, n. 1, p. 75–82, 2015.

DYE, C. et al. Data sharing in public health emergencies: a call to researchers. **Bull World Health Organ**, [S.l.], v. 94, n. 3, p. 158, 2016.

EDEN, K. B. et al. Barriers and facilitators to exchanging health information: a systematic review. **International journal of medical informatics**, [S.l.], v. 88, p. 44–51, 2016.

EHRSERVER. **Ehrserver**: the openehr clinical data repository. Last accessed: 2019-05-12, Available from: `https://cloudehrserver.com/`.

EKBLAW, A. C. **Medrec**: blockchain for medical data access, permission management and trend analysis. 2017. Tese (Doutorado em Ciência da Computação) — Massachusetts Institute of Technology, 2017.

EKBLAW, A. et al. A case study for blockchain in healthcare:"medrec" prototype for electronic health records and medical research data. In: IEEE OPEN & BIG DATA CONFERENCE, 2016. **Proceedings...** [S.l.: s.n.], 2016. v. 13, p. 13.

EMANI, S. et al. Patient perceptions of a personal health record: a test of the diffusion of innovation model. **Journal of medical Internet research**, [S.l.], v. 14, n. 6, 2012.

ERHARDT, K. D. et al. **Bismuth**: a blockchain-based program for verifying responsible data usage. 2017. Tese (Doutorado em Ciência da Computação) — Massachusetts Institute of Technology, 2017.

ESPOSITO, C.; CASTIGLIONE, A.; PALMIERI, F. Interoperable access control by means of a semantic approach. In: ADVANCED INFORMATION NETWORKING AND APPLICATIONS WORKSHOPS (WAINA), 2016 30TH INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 280–285.

FHIR. **Hl7/fhir**: fast healthcare interoperability resources. Last accessed: 2019-05-12, Available from: `http://hl7.org/fhir/documentation.html`.

FORD, E. W.; HESSE, B. W.; HUERTA, T. R. Personal health record use in the united states: forecasting future adoption levels. **Journal of medical Internet research**, [S.l.], v. 18, n. 3, 2016.

FRAGIDIS, L. L.; CHATZOGLOU, P. D. Development of nationwide electronic health record (nehr): an international survey. **Health Policy and Technology**, [S.l.], v. 6, n. 2, p. 124–133, 2017.

FRAGIDIS, L. L.; CHATZOGLOU, P. D. Implementation of a nationwide electronic health record (ehr) the international experience in 13 countries. **International journal of health care quality assurance**, [S.l.], v. 31, n. 2, p. 116–130, 2018.

FRICTON, J. R.; DAVIES, D. Personal health records to improve health information exchange and patient safety. In: **Advances in patient safety**: new directions and alternative approaches (vol. 4: technology and medication safety). [S.l.]: Agency for Healthcare Research and Quality (US), 2008.

FROST, J. H.; MASSAGLI, M. P. Social uses of personal health information within patientslikeme, an online patient community: what can happen when patients have access to one another's data. **Journal of medical Internet research**, [S.l.], v. 10, n. 3, 2008.

FUJI, K. T. et al. Standalone personal health records in the united states: meeting patient desires. **Health and technology**, [S.l.], v. 2, n. 3, p. 197–205, 2012.

GALSTER, M.; BUCHERER, E. A taxonomy for identifying and specifying non-functional requirements in service-oriented development. In: SERVICES-PART I, 2008. IEEE CONGRESS ON, 2008. **Anais...** [S.l.: s.n.], 2008. p. 345–352.

GARCÍA-BORGOÑON, L. et al. Software process modeling languages: a systematic literature review. **Information and Software Technology**, [S.l.], v. 56, n. 2, p. 103–116, 2014.

GARDIYAWASAM PUSSEWALAGE, H. S.; OLESHCHUK, V. A. A distributed multi-authority attribute based encryption scheme for secure sharing of personal health records. In: ACM ON SYMPOSIUM ON ACCESS CONTROL MODELS AND TECHNOLOGIES, 22., 2017. **Proceedings...** [S.l.: s.n.], 2017. p. 255–262.

GATE. **Gate**: general architecture for text engineering. department of computer science. Last accessed: 2019-05-12, Available from: `http://gate.ac.uk/`.

GEORGE, T. P.; HOPLA, D. L. Advantages of personal health records. **Nursing2015 Critical Care**, [S.l.], v. 10, n. 6, p. 10–12, 2015.

GOERTZEL, B. et al. **Singularitynet**: a decentralized, open market and inter-network for ais. 2017.

GOLDIM, J. R.; GIBBON, S. Between personal and relational privacy: understanding the work of informed consent in cancer genetics in brazil. **Journal of community genetics**, [S.l.], v. 6, n. 3, p. 287–293, 2015.

GONÇALVES-FERREIRA, D. et al. Openehr and general data protection regulation: evaluation of principles and requirements. **JMIR medical informatics**, [S.l.], v. 7, n. 1, p. e9845, 2019.

GOOSSEN, W.; GOOSSEN-BAREMANS, A.; VAN DER ZEL, M. Detailed clinical models: a review. **Healthcare informatics research**, [S.l.], v. 16, n. 4, p. 201–214, 2010.

GROENEN, C. J. et al. Improving maternity care using a personal health record: study protocol for a stepped-wedge, randomised, controlled trial. **Trials**, [S.l.], v. 17, n. 1, p. 202, 2016.

GUBBI, J. et al. Internet of things (iot): a vision, architectural elements, and future directions. **Future generation computer systems**, [S.l.], v. 29, n. 7, p. 1645–1660, 2013.

GUNTER, T. D.; TERRY, N. P. The emergence of national electronic health record architectures in the united states and australia: models, costs, and questions. **Journal of medical Internet research**, [S.l.], v. 7, n. 1, 2005.

GUTIéRREZ, P. P. **Cabolabs ehrserver**: service-oriented openehr repository for clinical data with composition commit, query and retrieve capabilities. 2018.

HAILEMICHAEL, M. A.; MARCO-RUIZ, L.; BELLIKA, J. G. Privacy-preserving statistical query and processing on distributed openehr data. **Studies in health technology and informatics**, [S.l.], v. 210, p. 766–770, 2015.

HAN, J. E. et al. Effect of electronic health record implementation in critical care on survival and medication errors. **The American journal of the medical sciences**, [S.l.], v. 351, n. 6, p. 576–581, 2016.

HAWIG, D. et al. Designing a distributed ledger technology system for interoperable and general data protection regulation–compliant health data exchange: a use case in blood glucose data. **Journal of Medical Internet Research**, [S.l.], v. 21, n. 6, p. e13665, 2019.

HE, C.; FAN, X.; LI, Y. Toward ubiquitous healthcare services with a novel efficient cloud platform. **IEEE Transactions on Biomedical Engineering**, [S.l.], v. 60, n. 1, p. 230–234, 2013.

HEART, T.; BEN-ASSULI, O.; SHABTAI, I. A review of phr, emr and ehr integration: a more personalized healthcare and public health policy. **Health Policy and Technology**, [S.l.], 2016.

HEART, T.; BEN-ASSULI, O.; SHABTAI, I. A review of phr, emr and ehr integration: a more personalized healthcare and public health policy. **Health Policy and Technology**, [S.l.], v. 6, n. 1, p. 20–25, 2017.

HEINTZMAN, N.; KLEINBERG, S. Using uncertain data from body-worn sensors to gain insight into type 1 diabetes. **Journal of biomedical informatics**, [S.l.], v. 63, p. 259–268, 2016.

HENZE, M. et al. Distributed configuration, authorization and management in the cloud-based internet of things. In: IEEE TRUSTCOM/BIGDATASE/ICESS, 2017., 2017. **Anais. . .** [S.l.: s.n.], 2017. p. 185–192.

HIMSS. **What is interoperability?** 2018.

HIPAA. **Health insurance portability and accountability act**. Last accessed: 2019-04-02, Available from: `https://www.hhs.gov/hipaa/`.

HOGAN, T. P. et al. Technology-assisted patient access to clinical information: an evaluation framework for blue button. **JMIR research protocols**, [S.l.], v. 3, n. 1, p. e18, 2014.

HORAN, T. A.; BOTTS, N. E.; BURKHARD, R. J. A multidimensional view of personal health systems for underserved populations. **Journal of medical Internet research**, [S.l.], v. 12, n. 3, 2010.

HU, H.; ELKUS, A.; KERSCHBERG, L. A personal health recommender system incorporating personal health records, modular ontologies, and crowd-sourced data. In: ADVANCES IN SOCIAL NETWORKS ANALYSIS AND MINING (ASONAM), 2016 IEEE/ACM INTERNATIONAL CONFERENCE ON, 2016. **Anais. . .** [S.l.: s.n.], 2016. p. 1027–1033.

HUDA, M. N.; YAMADA, S.; SONEHARA, N. Privacy-aware access to patient-controlled personal health records in emergency situations. In: PERVASIVE COMPUTING TECHNOLOGIES FOR HEALTHCARE, 2009. PERVASIVEHEALTH 2009. 3RD INTERNATIONAL CONFERENCE ON, 2009. **Anais. . .** [S.l.: s.n.], 2009. p. 1–6.

HUDSON, D. L.; COHEN, M. E. Uncertainty and complexity in personal health records. In: ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY (EMBC), 2010 ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, 2010. **Anais. . .** [S.l.: s.n.], 2010. p. 6773–6776.

HUSSEIN, S. E.; BADR, S. M. Healthcare cloud integration using distributed cloud storage and hybrid image compression. **International Journal of Computer Applications**, [S.l.], v. 80, n. 3, 2013.

ICD. **Family of international classifications**. Last accessed: 2019-04-02, Available from: `http://www.who.int/classifications/en/`.

ICHIKAWA, D.; KASHIYAMA, M.; UENO, T. Tamper-resistant mobile health using blockchain technology. **JMIR mHealth and uHealth**, [S.l.], v. 5, n. 7, 2017.

ICPC. **International classification of primary care**. Last accessed: 2019-04-02, Available from: `http://www.globalfamilydoctor.com/`.

IHE. **Integrating the healthcare enterprise**. Last accessed: 2019-04-02, Available from: `http://www.ihe.net/`.

IKKINK, H. K. **Gradle dependency management**. [S.l.]: Packt Publishing Ltd, 2015.

INFORMATICS, O. **Clinical knowledge manager**. 2018.

ISCO. **International standard classification of occupations**. [S.l.]: International Labour Organization, 2016.

ISLAM, S. R. et al. The internet of things for health care: a comprehensive survey. **IEEE Access**, [S.l.], v. 3, p. 678–708, 2015.

ISO. Health informatics - electronic health record - definition, scope and context. **Technical Report**, [S.l.], n. ISO/TR 20514:2005(E), p. 6–8, 2005.

ISO. Health informatics — personal health records — definition, scope and context. **Technical Report**, [S.l.], n. ISO/TR 14292:2012(E), p. 5–10, 2012.

ISO. Health informatics — capacity-based ehealth architecture roadmap — part 2: architectural components and maturity model. **Technical Report**, [S.l.], n. ISO/TR TR14639-2, 2014.

ISO. **Iso/tc 215 - health informatics**. Last accessed: 2019-04-02, Available from: `https://www.iso.org/committee/54960.html`.

ISO. Health informatics —- guidance on health information privacy education in healthcare organizations. **Technical Report**, [S.l.], n. ISO/TR 18638:2017, 2017.

ISTEPHAN, S.; SIADAT, M.-R. Unstructured medical image query using big data–an epilepsy case study. **Journal of biomedical informatics**, [S.l.], v. 59, p. 218–226, 2016.

JAIN, A. **Mastering apache storm**: real-time big data streaming using kafka, hbase and redis. [S.l.]: Packt Publishing Ltd, 2017.

JAMOOM, E.; YANG, N.; HING, E. Adoption of certified electronic health record systems and electronic information sharing in physician offices: united states, 2013 and 2014. **NCHS data brief**, [S.l.], n. 236, p. 1–8, 2016.

JOHNSON, A. E. et al. Mimic-iii, a freely accessible critical care database. **Scientific data**, [S.l.], v. 3, 2016.

JONES, D. A. et al. Characteristics of personal health records: findings of the medical library association/national library of medicine joint electronic personal health record task force. **Journal of the Medical Library Association: JMLA**, [S.l.], v. 98, n. 3, p. 243, 2010.

JV, I. D.; KALYANKAR, N.; KHAMITKAR, S. Computer network performance evaluation based on different data packet size using omnet++ simulation environment. **International Journal of Advanced Research in Computer Science and Technology**, [S.l.], 2014.

KAELBER, D. C. et al. A research agenda for personal health records (phrs). **Journal of the American Medical Informatics Association**, [S.l.], v. 15, n. 6, p. 729–736, 2008.

KAUR, K.; RANI, R. Managing data in healthcare information systems: many models, one solution. **Computer**, [S.l.], v. 48, n. 3, p. 52–59, 2015.

KEMKAR, O. S.; KALODE, P. Formulation of distributed electronic patient record (depr) system using openemr concept. **IJEIR**, [S.l.], v. 4, n. 1, p. 85–89, 2015.

KHARRAZI, H. et al. Mobile personal health records: an evaluation of features and functionality. **International journal of medical informatics**, [S.l.], v. 81, n. 9, p. 579–593, 2012.

KIM, E.-H. et al. Challenges to using an electronic personal health record by a low-income elderly population. **Journal of medical Internet research**, [S.l.], v. 11, n. 4, 2009.

KITCHENHAM, B. A.; CHARTERS, S. **Guidelines for performing systematic literature reviews in software engineering**. [S.l.]: Keele University, 2007. Technical Report. (EBSE-2007-01).

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. **Information and software technology**, [S.l.], v. 55, n. 12, p. 2049–2075, 2013.

KLEIN, D. M. et al. Use of the blue button online tool for sharing health information: qualitative interviews with patients and providers. **Journal of medical Internet research**, [S.l.], v. 17, n. 8, 2015.

KOTHARI, C. R. **Research methodology**: methods and techniques. [S.l.]: New Age International, 2004.

KRAAN, C. et al. Personal health records: solving barriers to enhance adoption. **University of Twente**, [S.l.], 2015.

KRAFT, D. Difficulty control for blockchain-based consensus systems. **Peer-to-Peer Networking and Applications**, [S.l.], v. 9, n. 2, p. 397–413, 2016.

KRASOWSKI, M. D. et al. Promoting improved utilization of laboratory testing through changes in an electronic medical record: experience at an academic medical center. **BMC medical informatics and decision making**, [S.l.], v. 15, n. 1, p. 11, 2015.

KSIMON, S.; SONAI MUTHU ANBANANTHEN, K.; LEE, S. A ubiquitous personal health record (uphr) framework. In: INTERNATIONAL CONFERENCE ON ADVANCED COMPUTER SCIENCE AND ELECTRONICS INFORMATION (ICACSEI 2013), 2013., 2013. **Anais...** [S.l.: s.n.], 2013.

KUHN, T. A survey and classification of controlled natural languages. **Computational Linguistics**, [S.l.], v. 40, n. 1, p. 121–170, 2014.

KUO, T.-T.; KIM, H.-E.; OHNO-MACHADO, L. Blockchain distributed ledger technologies for biomedical and health care applications. **Journal of the American Medical Informatics Association**, [S.l.], v. 24, n. 6, p. 1211–1220, 2017.

KYAZZE, M.; WESSON, J.; NAUDE, K. The design and implementation of a ubiquitous personal health record system for south africa. **Studies in health technology and informatics**, [S.l.], v. 206, p. 29, 2014.

LAFKY, D. B.; HORAN, T. A. Personal health records: consumer attitudes toward privacy and security of their personal health information. **Health Informatics Journal**, [S.l.], v. 17, n. 1, p. 63–71, 2011.

LAHTEENMAKI, J.; LEPPANEN, J.; KAIJANRANTA, H. Interoperability of personal health records. In: ENGINEERING IN MEDICINE AND BIOLOGY SOCIETY, 2009. EMBC 2009. ANNUAL INTERNATIONAL CONFERENCE OF THE IEEE, 2009. **Anais...** [S.l.: s.n.], 2009. p. 1726–1729.

LAURA, K.; KOIVUMÄKI, T.; SARANIEMI, S. D. Business models for platform operators in mydata based ecosystem–context preventive healthcare. **Marketing**, [S.l.], 2016.

LAZAROVICH, A. **Invisible ink**: blockchain for data privacy. 2015. Tese (Doutorado em Ciência da Computação) — Massachusetts Institute of Technology, 2015.

LEGAZ-GARCÍA, M. d. C. et al. Transformation of standardized clinical models based on owl technologies: from cem to openehr archetypes. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 3, p. 536–544, 2015.

LI, B. et al. Experiences of building a medical data acquisition system based on two-level modeling. **International Journal of Medical Informatics**, [S.l.], 2018.

LI, M. et al. Scalable and secure sharing of personal health records in cloud computing using attribute-based encryption. **IEEE transactions on parallel and distributed systems**, [S.l.], v. 24, n. 1, p. 131–143, 2013.

LI, S.; DA XU, L.; ZHAO, S. The internet of things: a survey. **Information Systems Frontiers**, [S.l.], v. 17, n. 2, p. 243–259, 2015.

LIU, J.; HUANG, X.; LIU, J. K. Secure sharing of personal health records in cloud computing: ciphertext-policy attribute-based signcryption. **Future Generation Computer Systems**, [S.l.], v. 52, p. 67–76, 2015.

LIU, L. S.; SHIH, P. C.; HAYES, G. R. Barriers to the adoption and use of personal health record systems. In: CONFERENCE, 2011., 2011. **Proceedings...** [S.l.: s.n.], 2011. p. 363–370.

LOINC. **Logical observation identifiers names and codes**. Last accessed: 2019-04-02, Available from: `https://loinc.org/`.

LOZANO-RUBÍ, R. et al. Ontocr: a cen/iso-13606 clinical repository based on ontologies. **Journal of biomedical informatics**, [S.l.], v. 60, p. 224–233, 2016.

LU, H. et al. Towards user-oriented rbac model. **Journal of Computer Security**, [S.l.], v. 23, n. 1, p. 107–129, 2015.

LUO, G.; TANG, C.; THOMAS, S. B. Intelligent personal health record: experience and open issues. **Journal of medical systems**, [S.l.], v. 36, n. 4, p. 2111–2128, 2012.

MA, M.; SHI, G.; LI, F. Privacy-oriented blockchain-based distributed key management architecture for hierarchical access control in the iot scenario. **IEEE Access**, [S.l.], v. 7, p. 34045–34059, 2019.

MAHER, M. et al. User-centered design groups to engage patients and caregivers with a personalized health information technology tool. **Biology of Blood and Marrow Transplantation**, [S.l.], v. 22, n. 2, p. 349–358, 2016.

MALIK, M.; SALEEM, M. Nlp-driven ontology modeling for handling event semantics in nl constraints. In: INNOVATIVE COMPUTING TECHNOLOGY (INTECH), 2016 SIXTH INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 485–490.

MANDEL, J. C. et al. Smart on fhir: a standards-based, interoperable apps platform for electronic health records. **Journal of the American Medical Informatics Association**, [S.l.], v. 23, n. 5, p. 899–908, 2016.

MARCOS, C. et al. Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on hl7's virtual medical record standard. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 3, p. 587–599, 2015.

MCCOY, A. B. et al. Matching identifiers in electronic health records: implications for duplicate records and patient safety. **BMJ quality & safety**, [S.l.], v. 22, n. 3, p. 219–224, 2013.

MCFARLANE, C. et al. Patientory: a healthcare peer-to-peer emr storage network v1. **Entrust Inc.: Addison, TX, USA**, [S.l.], 2017.

MCGHIN, T. et al. Blockchain in healthcare applications: research challenges and opportunities. **Journal of Network and Computer Applications**, [S.l.], 2019.

MEIER, C. A.; FITZGERALD, M. C.; SMITH, J. M. ehealth: extending, enhancing, and evolving health care. **Annual review of biomedical engineering**, [S.l.], v. 15, p. 359–382, 2013.

MENDES, D. et al. Ontology driven controlled natural language clinical decision support system for the cardiovascular specialty. **Procedia Technology**, [S.l.], v. 16, p. 1493–1501, 2014.

METTLER, M. Blockchain technology in healthcare: the revolution starts here. In: IEEE 18TH INTERNATIONAL CONFERENCE ON E-HEALTH NETWORKING, APPLICATIONS AND SERVICES (HEALTHCOM), 2016., 2016. **Anais...** [S.l.: s.n.], 2016. p. 1–3.

MIHAJLOVIĆ, V. et al. Wearable, wireless eeg solutions in daily life applications: what are we missing? **IEEE journal of biomedical and health informatics**, [S.l.], v. 19, n. 1, p. 6–21, 2015.

MILSTEIN, R.; BLANKART, C. R. The health care strengthening act: the next level of integrated care in germany. **Health Policy**, [S.l.], v. 120, n. 5, p. 445–451, 2016.

MO, H. et al. Desiderata for computable representations of electronic health records-driven phenotype algorithms. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 6, p. 1220–1230, 2015.

MOORHOUSE, J. et al. Modelling and simulation of peer-to-peer overlay network protocols using oversim. In: COMPUTER MODELLING AND SIMULATION (UKSIM), 2013 UKSIM 15TH INTERNATIONAL CONFERENCE ON, 2013. **Anais...** [S.l.: s.n.], 2013. p. 144–149.

MORABITO, R. A performance evaluation of container technologies on internet of things devices. In: COMPUTER COMMUNICATIONS WORKSHOPS (INFOCOM WKSHPS), 2016 IEEE CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 999–1000.

MORAES, J. L. C. de et al. A methodology based on openehr archetypes and software agents for developing e-health applications reusing legacy systems. **Computer methods and programs in biomedicine**, [S.l.], v. 134, p. 267–287, 2016.

MORENO-CONDE, A. et al. Clinical information modeling processes for semantic interoperability of electronic health records: systematic review and inductive analysis. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 4, p. 925–934, 2015.

MORGENTHALER, J. **Moving toward an open standard, universal health record**. [S.l.]: Smart Publications, 2011.

MORTENSEN, J. M. et al. Using the wisdom of the crowds to find critical errors in biomedical ontologies: a study of snomed ct. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 3, p. 640–648, 2015.

MUÑOZ, P. et al. The iso/en 13606 standard for the interoperable exchange of electronic health records. **Journal of Healthcare Engineering**, [S.l.], v. 2, n. 1, p. 1–24, 2011.

MXOLI, A.; GERBER, M.; MOSTERT-PHIPPS, N. Information security risk measures for cloud-based personal health records. In: INFORMATION SOCIETY (I-SOCIETY), 2014 INTERNATIONAL CONFERENCE ON, 2014. **Anais…** [S.l.: s.n.], 2014. p. 187–193.

NAKAMOTO, S. **Bitcoin**: a peer-to-peer electronic cash system. Last accessed: 2019-04-02, Available from: `https://bitcoin.org/bitcoin.pdf`.

NARAYANAN, A. et al. **Bitcoin and cryptocurrency technologies**: a comprehensive introduction. [S.l.]: Princeton University Press, 2016.

NAZI, K. M. The personal health record paradox: health care professionals' perspectives and the information ecology of personal health record systems in organizational and clinical settings. **Journal of medical Internet research**, [S.l.], v. 15, n. 4, 2013.

NAZI, K. M. et al. Embracing a health services research perspective on personal health records: lessons learned from the va my healthevet system. **Journal of general internal medicine**, [S.l.], v. 25, n. 1, p. 62–67, 2010.

NEO4J. **Neo4j**: open source graph database. graph dbms. Last accessed: 2019-06-14.

NEW, J. P. et al. Putting patients in control of data from electronic health records. **BMJ**, [S.l.], v. 360, p. j5554, 2018.

ODGERS, D. J.; DUMONTIER, M. Mining electronic health records using linked data. **AMIA Summits on Translational Science Proceedings**, [S.l.], v. 2015, p. 217, 2015.

OEMIG, F.; BLOBEL, B. Natural language processing supporting interoperability in healthcare. In: **Text mining**. [S.l.]: Springer, 2014. p. 137–156.

OPENEHR. **openehr - an open domain-driven platform for developing flexible e-health systems**. Last accessed: 2019-04-02, Available from: `http://www.openehr.org/`.

OPENID. **Final**: openid connect core 1.0 incorporating errata set 1. Last accessed: 2019-04-02, Available from: `http://openid.net/specs/openid-connect-core-1_0.html`.

OROM, H. et al. Relationships as medicine: quality of the physician–patient relationship determines physician influence on treatment recommendation adherence. **Health services research**, [S.l.], v. 53, n. 1, p. 580–596, 2018.

OUADDAH, A.; ABOU ELKALAM, A.; AIT OUAHMAN, A. Fairaccess: a new blockchain-based access control framework for the internet of things. **Security and Communication Networks**, [S.l.], v. 9, n. 18, p. 5943–5964, 2016.

OWL. **The owl api**: java api and reference implementation for owl ontologies. Last accessed: 2019-06-14, Available from: `http://owlapi.sourceforge.net/`.

OZOK, A. A. et al. Usability and perceived usefulness of personal health records for preventive health care: a case study focusing on patients' and primary care providers' perspectives. **Applied ergonomics**, [S.l.], v. 45, n. 3, p. 613–628, 2014.

O'DRISCOLL, A.; DAUGELAITE, J.; SLEATOR, R. D. 'big data', hadoop and cloud computing in genomics. **Journal of biomedical informatics**, [S.l.], v. 46, n. 5, p. 774–781, 2013.

PAIS, S.; PARRY, D.; HUANG, Y. Suitability of fast healthcare interoperability resources (fhir) for wellness data. In: HAWAII INTERNATIONAL CONFERENCE ON SYSTEM SCIENCES, 50., 2017. **Proceedings...** [S.l.: s.n.], 2017.

PANDIT, R. R.; BOLAND, M. V. Impact of digital imaging and communications in medicine workflow on the integration of patient demographics and ophthalmic test data. **Ophthalmology**, [S.l.], v. 122, n. 2, p. 227–232, 2015.

PARK, Y. R. et al. Is blockchain technology suitable for managing personal health records? mixed-methods study to test feasibility. **Journal of medical Internet research**, [S.l.], v. 21, n. 2, p. e12533, 2019.

PARSER. **The stanford parser**: a statistical parser. Last accessed: 2019-06-14.

PATEL, C. et al. Trialx: using semantic technologies to match patients to relevant clinical trials based on their personal health records. **Web Semantics: Science, Services and Agents on the World Wide Web**, [S.l.], v. 8, n. 4, p. 342–347, 2010.

PELEG, M. et al. Mobiguide: a personalized and patient-centric decision-support system and its evaluation in the atrial fibrillation and gestational diabetes domains. **User Modeling and User-Adapted Interaction**, [S.l.], v. 27, n. 2, p. 159–213, 2017.

PETTICREW, M.; ROBERTS, H. **Systematic reviews in the social sciences**: a practical guide. [S.l.]: John Wiley & Sons, 2008.

POSTGRESQL. **Postgresql**: open-source database. Last accessed: 2019-06-14.

POWERS, D. M. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. **Journal of Machine Learning Technologies**, [S.l.], v. 2, n. 1, p. 37–63, 2011.

PRICE, M. et al. Conditions potentially sensitive to a personal health record (phr) intervention, a systematic review. **BMC medical informatics and decision making**, [S.l.], v. 15, n. 1, p. 32, 2015.

PUTHAL, D. et al. The blockchain as a decentralized security framework [future directions]. **IEEE Consumer Electronics Magazine**, [S.l.], v. 7, n. 2, p. 18–21, 2018.

QIU, D. et al. Regression testing of web service: a systematic mapping study. **ACM Computing Surveys (CSUR)**, [S.l.], v. 47, n. 2, p. 21, 2015.

RAJKOMAR, A. et al. Scalable and accurate deep learning with electronic health records. **NPJ Digital Medicine**, [S.l.], v. 1, n. 1, p. 18, 2018.

RANDALL, D.; GOEL, P.; ABUJAMRA, R. Blockchain applications and use cases in health information technology. **J Health Med Informat**, [S.l.], v. 8, n. 276, p. 2, 2017.

RATHERT, C. et al. Seven years after meaningful use: physicians' and nurses' experiences with electronic health records. **Health care management review**, [S.l.], v. 44, n. 1, p. 30–40, 2019.

REEDER, B.; DAVID, A. Health at hand: a systematic review of smart watch uses for health and wellness. **Journal of Biomedical Informatics**, [S.l.], v. 63, p. 269–276, 2016.

RETI, S. R. et al. Improving personal health records for patient-centered care. **Journal of the American Medical Informatics Association**, [S.l.], v. 17, n. 2, p. 192–195, 2010.

ROEHRS, A.; COSTA, C. A. da; ROSA RIGHI, R. da. Omniphr: a distributed architecture model to integrate personal health records. **Journal of biomedical informatics**, [S.l.], v. 71, p. 70–81, 2017.

ROEHRS, A. et al. Personal health records: a systematic literature review. **Journal of Medical Internet Research**, [S.l.], v. 19, n. 1, p. e13, 2017.

ROEHRS, A. et al. Toward a model for personal health records interoperability. **IEEE Journal of Biomedical and Health Informatics**, [S.l.], 2018.

ROEHRS, A. et al. Analyzing the performance of a blockchain-based personal health record implementation. **Journal of biomedical informatics**, [S.l.], p. 103140, 2019.

RONO, D. K. **A restful e-health interoperability platform**: case of nairobi county health facilities. 2016. Tese (Doutorado em Ciência da Computação) — Strathmore University, 2016.

ROSA RIGHI, R. da et al. Autoelastic: automatic resource elasticity for high performance applications in the cloud. **IEEE Transactions on Cloud Computing**, [S.l.], v. 4, n. 1, p. 6–19, 2016.

ROUMIA, M.; STEINHUBL, S. Improving cardiovascular outcomes using electronic health records. **Current cardiology reports**, [S.l.], v. 16, n. 2, p. 1–6, 2014.

SACHDEVA, S.; BATRA, S.; BHALLA, S. Evolving large scale healthcare applications using open standards. **Health Policy and Technology**, [S.l.], v. 6, n. 4, p. 410–425, 2017.

SAFAVI, S.; SHUKUR, Z. Conceptual privacy framework for health information on wearable device. **PloS one**, [S.l.], v. 9, n. 12, p. e114306, 2014.

SARIPALLE, R.; RUNYAN, C.; RUSSELL, M. Using hl7 fhir to achieve interoperability in patient health record. **Journal of biomedical informatics**, [S.l.], p. 103188, 2019.

SAVOVA, G. K. et al. Mayo clinical text analysis and knowledge extraction system (ctakes): architecture, component evaluation and applications. **Journal of the American Medical Informatics Association**, [S.l.], v. 17, n. 5, p. 507–513, 2010.

SCHINASI, L. H. et al. Using electronic health record data for environmental and place based population health research: a systematic review. **Annals of epidemiology**, [S.l.], v. 28, n. 7, p. 493–502, 2018.

SEÑOR, I. C.; FERNÁNDEZ-ALEMÁN, J. L.; TOVAL, A. Are personal health records safe? a review of free web-accessible personal health record privacy policies. **Journal of Medical Internet Research**, [S.l.], v. 14, n. 4, 2012.

SHAH, G. H. et al. Characteristics of local health departments associated with implementation of electronic health records and other informatics systems. **Public Health Reports**, [S.l.], v. 131, n. 2, p. 272–282, 2016.

SHARMA, P. K.; CHEN, M.-Y.; PARK, J. H. A software defined fog node based distributed blockchain cloud architecture for iot. **IEEE Access**, [S.l.], v. 6, p. 115–124, 2017.

SHARMA, V. et al. Secure and energy-efficient handover in fog networks using blockchain-based dmm. **IEEE Communications Magazine**, [S.l.], v. 56, n. 5, p. 22–31, 2018.

SHE, W. et al. Homomorphic consortium blockchain for smart home system sensitive data privacy preserving. **IEEE Access**, [S.l.], v. 7, p. 62058–62070, 2019.

SHIMADA, S. L. et al. Personal health record reach in the veterans health administration: a cross-sectional analysis. **Journal of medical Internet research**, [S.l.], v. 16, n. 12, p. e272, 2014.

SHORTLIFFE, E. H.; CIMINO, J. J. **Biomedical informatics**: computer applications in health care and biomedicine. [S.l.]: Springer Science & Business Media, 2013.

SHOWELL, C. Barriers to the use of personal health records by patients: a structured review. **PeerJ**, [S.l.], v. 5, p. e3268, 2017.

SIEK, K. A. et al. Designing a personal health application for older adults to manage medications: a comprehensive case study. **Journal of medical systems**, [S.l.], v. 35, n. 5, p. 1099–1121, 2011.

SIMPSON, K. R. Electronic health records. **MCN: The American Journal of Maternal/Child Nursing**, [S.l.], v. 40, n. 1, p. 68, 2015.

SNOMED. **Systematized nomenclature of medicine**. Last accessed: 2019-04-02, Available from: `http://www.snomed.org/snomed-ct`.

SOCEANU, I. A. Managing the interoperability and privacy of e-health systems as an interdisciplinary challenge. **SYSTEMICS, CYBERNETICS AND INFORMATICS**, [S.l.], v. 14, n. 5, 2016.

SOHAIB, O. et al. Integrating design thinking into extreme programming. **Journal of Ambient Intelligence and Humanized Computing**, [S.l.], v. 10, n. 6, p. 2485–2492, 2019.

SONG, Y.-T. et al. Standard-based patient-centered personal health record system. In: INTERNATIONAL CONFERENCE ON UBIQUITOUS INFORMATION MANAGEMENT AND COMMUNICATION, 11., 2017. **Proceedings...** [S.l.: s.n.], 2017. p. 63.

SPIL, T.; KLEIN, R. Personal health records success: why google health failed and what does that mean for microsoft healthvault? In: SYSTEM SCIENCES (HICSS), 2014 47TH HAWAII INTERNATIONAL CONFERENCE ON, 2014. **Anais...** [S.l.: s.n.], 2014. p. 2818–2827.

SPIL, T.; KLEIN, R. The personal health future. **Health policy and technology**, [S.l.], v. 4, n. 2, p. 131–136, 2015.

STAGNARO, C. White paper: innovative blockchain uses in health care. **Freed Associates**, [S.l.], 2017.

STARDOG. **Stardog**: enterprise knowledge graph database. semantic graph dbms. Last accessed: 2019-06-14.

STEELE, R.; MIN, K.; LO, A. Personal health record architectures: technology infrastructure implications and dependencies. **Journal of the Association for Information Science and Technology**, [S.l.], v. 63, n. 6, p. 1079–1091, 2012.

STOICA, I. et al. Chord: a scalable peer-to-peer lookup service for internet applications. **ACM SIGCOMM Computer Communication Review**, [S.l.], v. 31, n. 4, p. 149–160, 2001.

SUJANSKY, W.; KUNZ, D. A standard-based model for the sharing of patient-generated health information with electronic health records. **Personal and Ubiquitous Computing**, [S.l.], v. 19, n. 1, p. 9–25, 2015.

SUNYAEV, A.; CHORNYI, D. Supporting chronic disease care quality: design and implementation of a health service and its integration with electronic health records. **Journal of Data and Information Quality (JDIQ)**, [S.l.], v. 3, n. 2, p. 3, 2012.

SZABO, N. Smart contracts: building blocks for digital markets. **EXTROPY: The Journal of Transhumanist Thought,(16)**, [S.l.], v. 18, p. 2, 1996.

TA, V.-D.; LIU, C.-M.; NKABINDE, G. W. Big data stream computing in healthcare real-time analytics. In: CLOUD COMPUTING AND BIG DATA ANALYSIS (ICCCBDA), 2016 IEEE INTERNATIONAL CONFERENCE ON, 2016. **Anais...** [S.l.: s.n.], 2016. p. 37–42.

TANG, P. C. et al. Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption. **Journal of the American Medical Informatics Association**, [S.l.], v. 13, n. 2, p. 121–126, 2006.

TANG, P. C.; LANSKY, D. The missing link: bridging the patient–provider health information gap. **Health Affairs**, [S.l.], v. 24, n. 5, p. 1290–1295, 2005.

TAPSCOTT, D.; TAPSCOTT, A. **Blockchain revolution**: how the technology behind bitcoin is changing money, business, and the world. [S.l.]: Penguin, 2016.

TEODORO, D. et al. Orbda: an openehr benchmark dataset for performance assessment of electronic health record servers. **PloS one**, [S.l.], v. 13, n. 1, p. e0190028, 2018.

THAKKAR, P.; NATHAN, S.; VISWANATHAN, B. Performance benchmarking and optimizing hyperledger fabric blockchain platform. In: IEEE 26TH INTERNATIONAL SYMPOSIUM ON MODELING, ANALYSIS, AND SIMULATION OF COMPUTER AND TELECOMMUNICATION SYSTEMS (MASCOTS), 2018., 2018. **Anais…** [S.l.: s.n.], 2018. p. 264–276.

TONG, Y. et al. Cloud-assisted mobile-access of health data with privacy and auditability. **IEEE Journal of biomedical and health Informatics**, [S.l.], v. 18, n. 2, p. 419–429, 2014.

TSENG, S.-S. et al. Cbr-based negotiation rbac model for enhancing ubiquitous resources management. **International Journal of Information Management**, [S.l.], v. 37, n. 1, p. 1539–1550, 2017.

ULRIKSEN, G.-H.; PEDERSEN, R.; ELLINGSEN, G. Infrastructuring in healthcare through the openehr architecture. **Computer Supported Cooperative Work (CSCW)**, [S.l.], v. 26, n. 1-2, p. 33–69, 2017.

VAN GORP, P.; COMUZZI, M. Lifelong personal health data and application software via virtual machines in the cloud. **IEEE journal of biomedical and health informatics**, [S.l.], v. 18, n. 1, p. 36–45, 2014.

VATSAVAI, V.; SURAVARAPU, S.; MIR, N. F. Implementation of p2p file sharing using bi-directional chord protocol algorithm. In: **Advanced computer and communication engineering technology**. [S.l.]: Springer, 2016. p. 51–62.

VIRTUOSO. **Openlink virtuoso open-source edition.** Last accessed: 2019-06-14.

VO, H. T. et al. Internet of blockchains: techniques and challenges ahead. In: IEEE INTERNATIONAL CONFERENCE ON INTERNET OF THINGS (ITHINGS) AND IEEE GREEN COMPUTING AND COMMUNICATIONS (GREENCOM) AND IEEE CYBER, PHYSICAL AND SOCIAL COMPUTING (CPSCOM) AND IEEE SMART DATA (SMARTDATA), 2018., 2018. **Anais…** [S.l.: s.n.], 2018. p. 1574–1581.

VUOKKO, R. et al. Impacts of structuring the electronic health record: results of a systematic literature review from the perspective of secondary use of patient data. **International journal of medical informatics**, [S.l.], v. 97, p. 293–303, 2017.

WALPORT, M. Distributed ledger technology: beyond blockchain. **UK Government Office for Science**, [S.l.], 2016.

WANG, G. et al. Building a replicated logging system with apache kafka. **Proceedings of the VLDB Endowment**, [S.l.], v. 8, n. 12, p. 1654–1655, 2015.

WANG, K.; ZHANG, Z.; KIM, H. S. Reviewchain: smart contract based review system with multi-blockchain gateway. In: IEEE INTERNATIONAL CONFERENCE ON INTERNET OF THINGS (ITHINGS) AND IEEE GREEN COMPUTING AND COMMUNICATIONS (GREENCOM) AND IEEE CYBER, PHYSICAL AND SOCIAL COMPUTING (CPSCOM) AND IEEE SMART DATA (SMARTDATA), 2018., 2018. **Anais…** [S.l.: s.n.], 2018. p. 1521–1526.

WANG, T.; DOLEZEL, D. Usability of web-based personal health records: an analysis of consumers' perspectives. **Perspectives in Health Information Management**, [S.l.], v. 13, n. Spring, 2016.

WANG, Y.; KUNG, L.; BYRD, T. A. Big data analytics: understanding its capabilities and potential benefits for healthcare organizations. **Technological Forecasting and Social Change**, [S.l.], v. 126, p. 3–13, 2018.

WASS, S.; VIMARLUND, V. Same, same but different: perceptions of patients' online access to electronic health records among healthcare professionals. **Health informatics journal**, [S.l.], p. 1460458218779101, 2018.

WELLS, S. et al. Organizational strategies for promoting patient and provider uptake of personal health records. **Journal of the American Medical Informatics Association**, [S.l.], v. 22, n. 1, p. 213–222, 2014.

WEN, K.-Y. et al. Consumers' perceptions about and use of the internet for personal health records and health information exchange: analysis of the 2007 health information national trends survey. **Journal of medical Internet research**, [S.l.], v. 12, n. 4, 2010.

WILLIAMS, J. Social networking applications in health care: threats to the privacy and security of health information. In: ICSE WORKSHOP ON SOFTWARE ENGINEERING IN HEALTH CARE, 2010., 2010. **Proceedings. . .** [S.l.: s.n.], 2010. p. 39–49.

WOODS, S. S. et al. Patient experiences with full electronic access to health records and clinical notes through the my healthevet personal health record pilot: qualitative study. **Journal of medical Internet research**, [S.l.], v. 15, n. 3, 2013.

WOUNGANG, I. et al. Mr-chord: improved chord lookup performance in structured mobile p2p networks. **Systems Journal, IEEE**, [S.l.], v. 9, n. 3, p. 743–751, 2015.

WULFF, A. et al. An interoperable clinical decision-support system for early detection of sirs in pediatric intensive care using openehr. **Artificial intelligence in medicine**, [S.l.], v. 89, p. 10–23, 2018.

WYNIA, M.; DUNN, K. Dreams and nightmares: practical and ethical issues for patients and physicians using personal health records. **The Journal of Law, Medicine & Ethics**, [S.l.], v. 38, n. 1, p. 64–73, 2010.

XIA, C.; SONG, S. Resource allocation in hierarchical distributed ehr system based on improved poly-particle swarm. In: BIOMEDICAL ENGINEERING AND INFORMATICS (BMEI), 2012 5TH INTERNATIONAL CONFERENCE ON, 2012. **Anais. . .** [S.l.: s.n.], 2012. p. 1112–1116.

XIA, Q. et al. Medshare: trust-less medical data sharing among cloud service providers via blockchain. **IEEE Access**, [S.l.], v. 5, p. 14757–14767, 2017.

YADAV, P. et al. Mining electronic health records (ehrs): a survey. **ACM Comput. Surv.**, New York, NY, USA, v. 50, n. 6, p. 85:1–85:40, Jan. 2018.

YANG, L.; HUANG, X.; LI, J. Discovering clinical information models online to promote interoperability of electronic health records: a feasibility study of openehr. **Journal of medical Internet research**, [S.l.], v. 21, n. 5, p. e13504, 2019.

YAO, X. et al. Privacy-preserving search over encrypted personal health record in multi-source cloud. **IEEE Access**, [S.l.], v. 6, p. 3809–3823, 2018.

YUE, X. et al. Healthcare data gateways: found healthcare intelligence on blockchain with novel privacy risk control. **Journal of medical systems**, [S.l.], v. 40, n. 10, p. 218, 2016.

ZAHARIA, M. et al. Apache spark: a unified engine for big data processing. **Communications of the ACM**, [S.l.], v. 59, n. 11, p. 56–65, 2016.

ZHANG, P. et al. Metrics for assessing blockchain-based healthcare decentralized apps. In: IEEE 19TH INTERNATIONAL CONFERENCE ON E-HEALTH NETWORKING, APPLICATIONS AND SERVICES (HEALTHCOM), 2017., 2017. **Anais...** [S.l.: s.n.], 2017. p. 1–4.

ZHANG, P. et al. Fhirchain: applying blockchain to securely and scalably share clinical data. **Computational and structural biotechnology journal**, [S.l.], v. 16, p. 267–278, 2018.

ZULMAN, D. M. et al. Patient interest in sharing personal health record informationa web-based survey. **Annals of internal medicine**, [S.l.], v. 155, n. 12, p. 805–810, 2011.

APPENDIX A RESEARCH PORTALS

Selected research portals.

Table 22 – Selected research portals.

Acronym	**Portal Name**
ACM	ACM Digital Library
CiteSeerX	CiteSeerX Library
Google Scholar	Google Scholar
IEEE	IEEE Xplore Digital Library
IET	IET Digital Library
JMIR	JMIR Publications Library
PubMed	National Center for Biotechnology Information, US Nat. Lib. of Medicine
SciELO	Scientific Electronic Library Online
ScienceDirect	Elsevier B. V. ScienceDirect
Springer	Springer Science
Web of Science	Web of Science
Wiley	Wiley Online Library

Source: Prepared by the author

APPENDIX B EDITORS

List of editors.

Figure 30 – List of editors.

Publisher and Publications	
American College of Physicians	
- Annals of Internal Medicine	
Association for Computing Machinery	
- iConference	- Journal of Data and Information Quality
- International Conference on Software Engineering in Health Care	
BioMed Central	
- BMC Medical Informatics and Decision Making	
Elsevier	
- Applied Ergonomics	- International Journal of Medical Informatics
- Future Generation Computer Systems	- Journal of Biomedical Informatics
- Health Policy and Technology	- Web Semantics
IEEE	
- Transactions on Information Technology in Biomedicine	
- Transactions on Parallel and Distributed Systems	
- International Conference on System Sciences	
- International Conference on Pervasive Computing Technologies for Healthcare	
- International Conference of the Engineering in Medicine and Biology	
JMIR Publications	
- Journal of Medical Internet Research	
Medical Library Association	
- Journal of the Medical Library Association	
SAGE Publications	
- Health Informatics Journal	- Human Factors
Springer Science	
- Health and Technology	- Journal of Medical Systems
- Journal of General Internal Medicine	- Journal of Neuro-oncology
The Oxford University Press	
- Journal of the American Medical Informatics Association	
Wiley Online Library	
- The Journal of Law, Medicine & Ethics	
- Journal of the American Society for Information Science and Technology	

Source: Prepared by the author.

APPENDIX C USERS AND PROFILES

List of users and profiles access.

Figure 31 – List of users and profiles.

User or profile and articles	Providers and health care professionals			Laity		Public
	Physician, doctor	Nurse	Administrative, other	Patient or consumer	Relatives, caregiver, payer	Government, health plan
A01	✓	✓		✓	✓	✓
A02	✓			✓	✓	✓
A03	✓	✓	✓	✓	✓	✓
A04	✓	✓	✓	✓	✓	✓
A05			✓	✓		✓
A06		✓	✓	✓		✓
A07	✓	✓		✓		✓
A08	✓	✓	✓	✓	✓	✓
A09	✓	✓		✓	✓	✓
A10	✓				✓	✓
A11				✓		✓
A12	✓			✓		✓
A13	✓			✓		✓
A14	✓			✓		✓
A15	✓	✓		✓		✓
A16	✓	✓	✓	✓	✓	✓
A17	✓		✓	✓	✓	✓
A18	✓	✓	✓	✓	✓	✓
A19	✓			✓	✓	✓
A20	✓	✓		✓		✓
A21	✓	✓		✓	✓	✓
A22	✓	✓		✓	✓	✓
A23		✓		✓		✓
A24	✓		✓	✓		✓
A25	✓		✓	✓	✓	✓
A26	✓	✓	✓	✓	✓	
A27	✓			✓		✓
A28	✓	✓		✓	✓	✓
A29	✓	✓		✓	✓	✓
A30	✓			✓	✓	✓
A31	✓			✓	✓	✓
A32				✓	✓	✓
A33	✓	✓		✓	✓	✓
A34	✓	✓	✓	✓		
A35	✓	✓		✓	✓	✓
A36	✓			✓	✓	✓
A37	✓		✓	✓	✓	✓
A38	✓	✓		✓	✓	
A39	✓			✓	✓	
A40	✓			✓		✓
A41	✓			✓	✓	✓
A42	✓			✓	✓	✓
A43	✓			✓		✓
A44	✓	✓	✓	✓	✓	✓
A45	✓		✓	✓	✓	
A46	✓			✓	✓	✓
A47				✓	✓	✓
A48	✓		✓	✓	✓	✓

Source: Prepared by the author.